REMARKS

Claims 20-39 are pending in the application. Claims 22-28 and 31-39 are withdrawn as being drawn to non-elected inventions. Claims 20, 21, 29, and 30 are under consideration. Claims 20, 31, 34, and 37 have been amended to further clarify the intended subject matter of the claimed invention. Entry of these amendments is respectfully requested. Applicants reserve the right to prosecute non-elected subject matter in subsequent divisional applications.

Comments regarding restriction requirement

Applicants affirm the election with traverse of claims 20, 21, 29, and 30, corresponding to the invention of Group I, drawn to polypeptides, and the election of SEQ ID NO:1. The Examiner is reminded that claims 31, 32, 35, 38, and 39 are "method of use" claims which depend from the elected polypeptide product claims. Therefore, upon allowance of the polypeptide product claims, it is believed that claims 31, 32, 35, 38, and 39 should be rejoined and considered, in accordance with the Commissioner's Notice in the Official Gazette of March 26, entitled "Guidance on Treatment of Product and Process Claims in light of *In re Ochiai, In re Brouwer* and 35 U.S.C.§ 103(b)."

Objection to the claims

Claim 20 has been amended to address the objection to this claim. Withdrawal of the objection is therefore requested.

Rejection under 35 U.S.C. § 112, second paragraph

Claim 20 has been rejected under 35 U.S.C. § 112, second paragraph, as allegedly being indefinite in the recitation of the term "biologically active" on the grounds that "it is unclear what the scope of activities that is encompassed by this term includes" (Office Action, page 5). To expedite prosecution, claim 20 has been amended as follows:

20. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence of SEQ ID NO:1,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to an amino acid sequence of SEQ ID NO:1, wherein said polypeptide has endooligopeptidase activity, and

c) a biologically active fragment of a polypeptide having an amino acid sequence of SEQ ID NO:1, wherein said fragment has endooligopeptidase activity.

Applicants are amending the claim solely to obtain expeditious allowance of the instant application. Support for this amendment to claim 20 can be found in the specification, for example, at page 16, lines 14-19, and in Figure 5 which points out regions of homology between SEQ ID NO:1 and endooligopeptidase A (g2827886), and at page 51, line 26 through page 52, line 10, which describes assays for measuring protease activity. By this amendment, Applicants expressly do not disclaim equivalents of the invention which could include polypeptides or fragments having biological activities other than endooligopeptidase activity. The scope of the claim is now clear. Therefore, withdrawal of the rejection under U.S.C. § 112, second paragraph is respectfully requested.

Written description rejections under 35 U.S.C. § 112, first paragraph

Claims 20 and 29 have been rejected under the first paragraph of 35 U.S.C. 112 for alleged lack of an adequate written description. This rejection is respectfully traversed.

The requirements necessary to fulfill the written description requirement of 35 U.S.C. 112, first paragraph, are well established by case law.

> . . . the applicant must also convey with reasonable clarity to those skilled in the art that, as of the filing date sought, he or she was in possession *of the invention.* The invention is, for purposes of the "written description" inquiry, *whatever is now claimed. Vas-Cath, Inc. v. Mahurkar*, 19 USPQ2d 1111, 1117 (Fed. Cir. 1991)

Attention is also drawn to the Patent and Trademark Office's own "Guidelines for Examination of Patent Applications Under the 35 U.S.C. Sec. 112, para. 1", published January 5, 2001, which provide that :

> An applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics[42] which provide evidence that applicant was in possession of the claimed invention,[43] i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such

> characteristics.[44] What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail.[45] If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met.[46]

Thus, the written description standard is fulfilled by both what is specifically disclosed and what is conventional or well known to one skilled in the art.

SEQ ID NO:1 and SEQ ID NO:2 are specifically disclosed in the application (see, for example, page 16, lines 2-18). Variants of SEQ ID NO:1 are described, for example, at page 18, lines 15-18. Incyte clones in which the nucleic acids encoding the human protease HPRAP-1 were first identified and libraries from which those clones were isolated are described, for example, at page18, lines 2-8 of the Specification. Chemical and structural features of HPRAP-1 are described, for example, on page 18, lines 9-18. Given SEQ ID NO:1, one of ordinary skill in the art would recognize naturally-occurring variants of SEQ ID NO:1 having 95% sequence identity to SEQ ID NO:1. Accordingly, the Specification provides an adequate written description of the recited polypeptide sequences.

A. The Specification provides an adequate written description of the claimed variants and fragments of SEQ ID NO:1.

The Office Action has further asserted that the claims are not supported by an adequate written description because:

> No information beyond the characterization of SEQ ID NO:1 has been provided by applicants which would indicate that they had possession of the claimed genus of polypeptides (Office Action, page 6).

Such a position is believed to present a misapplication of the law.

The USPTO's own training materials on the application of the Written Description Guidelines to protein variants permit claims to a genus of variants without requiring that the actual amino acid sequence of every variant be disclosed. A reference sequence is considered to be "representative of the genus" when all members have a specified percentage identity and specified activity that defines the

"common attributes possessed by members of the genus" (see Revised Interim Written Description Guidelines Training Materials, "Synopsis of Application of Written Description Guidelines," Example 14: Product by Function; http://www.uspto.gov/web/patents/guides.htm).

1. The present claims specifically define the claimed genus through the recitation of chemical structure

Court cases in which "DNA claims" have been at issue (which are hence relevant to claims to proteins encoded by the DNA) commonly emphasize that the recitation of structural features or chemical or physical properties are important factors to consider in a written description analysis of such claims. For example, in *Fiers v. Revel*, 25 USPQ2d 1601, 1606 (Fed. Cir. 1993), the court stated that:

> If a conception of a DNA requires a precise definition, such as by structure, formula, chemical name or physical properties, as we have held, then a description also requires that degree of specificity.

In a number of instances in which claims to DNA have been found invalid, the courts have noted that the claims attempted to define the claimed DNA in terms of functional characteristics without any reference to structural features. As set forth by the court in *University of California v. Eli Lilly and Co.*, 43 USPQ2d 1398, 1406 (Fed. Cir. 1997):

> In claims to genetic material, however, a generic statement such as "vertebrate insulin cDNA" or "mammalian insulin cDNA," without more, is not an adequate written description of the genus because it does not distinguish the claimed genus from others, except by function.

Thus, the mere recitation of functional characteristics of a DNA, without the definition of structural features, has been a common basis by which courts have found invalid claims to DNA. For example, in *Lilly*, 43 USPQ2d at 1407, the court found invalid for violation of the written description requirement the following claim of U.S. Patent No. 4,652,525:

> 1. A recombinant plasmid replicable in procaryotic host containing within its nucleotide sequence a subsequence having the structure of the reverse transcript of an mRNA of a vertebrate, which mRNA encodes insulin.

In *Fiers*, 25 USPQ2d at 1603, the parties were in an interference involving the following count:

> A DNA which consists essentially of a DNA which codes for a human fibroblast interferon-beta polypeptide.

Party Revel in the *Fiers* case argued that its foreign priority application contained an adequate written description of the DNA of the count because that application mentioned a potential method for isolating the DNA. The Revel priority application, however, did not have a description of any particular DNA structure corresponding to the DNA of the count. The court therefore found that the Revel priority application lacked an adequate written description of the subject matter of the count.

Thus, in *Lilly* and *Fiers*, nucleic acids were defined on the basis of functional characteristics and were found not to comply with the written description requirement of 35 U.S.C. §112; *i.e.*, "an mRNA of a vertebrate, which mRNA encodes insulin" in *Lilly*, and "DNA which codes for a human fibroblast interferon-beta polypeptide" in *Fiers*. In contrast to the situation in *Lilly* and *Fiers*, the claims at issue in the present application define polypeptides in terms of chemical structure, rather than on functional characteristics. For example, the "variant language" of independent claim 20 recites chemical structure to define the claimed genus:

> 20. An isolated polypeptide selected from the group consisting of:...
> b) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to an amino acid sequence of SEQ ID NO:1, wherein said polypeptide has endooligopeptidase activity...

From the above it should be apparent that the claims of the subject application are fundamentally different from those found invalid in *Lilly* and *Fiers*. The subject matter of the present claims is defined in terms of the chemical structure of SEQ ID NO:1. In the present case, there is no reliance merely on a description of functional characteristics of the [polynucleotides or polypeptides or antibodies which specifically bind to the polypeptides, *as the case may be*] recited by the claims. In fact, there is no recitation of functional characteristics. Moreover, if such functional recitations were included, it would add to the structural characterization of the recited [polynucleotides or polypeptides or antibodies which specifically bind to the polypeptides, *as the case may be*]. The [polynucleotides or polypeptides or antibodies which specifically bind to the polypeptides, *as the case may be*] defined in the claims of the present application recite structural features, and cases such as *Lilly* and *Fiers* stress

that the recitation of structure is an important factor to consider in a written description analysis of claims of this type. By failing to base its written description inquiry "on whatever is now claimed," the Office Action failed to provide an appropriate analysis of the present claims and how they differ from those found not to satisfy the written description requirement in *Lilly* and *Fiers*

2. The present claims do not define a genus which is "highly variant"

Furthermore, the claims at issue do not describe a genus which could be characterized as "highly variant." Available evidence illustrates that the claimed genus is of narrow scope.

In support of this assertion, the Examiner's attention is directed to the enclosed reference by Brenner et al. ("Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships," Proc. Natl. Acad. Sci. USA (1998) 95:6073-6078). Through exhaustive analysis of a data set of proteins with known structural and functional relationships and with <90% overall sequence identity, Brenner et al. have determined that 30% identity is a reliable threshold for establishing evolutionary homology between two sequences aligned over at least 150 residues. (Brenner et al., pages 6073 and 6076.) Furthermore, local identity is particularly important in this case for assessing the significance of the alignments, as Brenner et al. further report that ≥40% identity over at least 70 residues is reliable in signifying homology between proteins. (Brenner et al., page 6076.)

The present application is directed, *inter alia*, to human protease proteins related to the amino acid sequence of SEQ ID NO:1. In accordance with Brenner et al, naturally occurring molecules may exist which could be characterized as human protease proteins and which have as little as 40% identity over at least 70 residues to SEQ ID NO:1. The "variant language" of the present claims recites, for example, polypeptides encoding "a naturally-occurring amino acid sequence having at least 95% sequence identity to the sequence of SEQ ID NO:1" (note that SEQ ID NO:1 has 459 amino acid residues). This variation is far less than that of all potential human protease proteins related to SEQ ID NO:1, i.e., those human protease proteins having as little as 40% identity over at least 70 residues to SEQ ID NO:1.

3. The state of the art at the time of the present invention is further advanced than at the time of the *Lilly* and *Fiers* applications

In the *Lilly* case, claims of U.S. Patent No. 4,652,525 were found invalid for failing to comply with the written description requirement of 35 U.S.C. §112. The '525 patent claimed the benefit of priority of two applications, Application Serial No. 801,343 filed May 27, 1977, and Application Serial No. 805,023 filed June 9, 1977. In the *Fiers* case, party Revel claimed the benefit of priority of an Israeli application filed on November 21, 1979. Thus, the written description inquiry in those case was based on the state of the art at essentially at the "dark ages" of recombinant DNA technology.

The present application has a priority date of May 1, 1998. Much has happened in the development of recombinant DNA technology in the 19 or more years from the time of filing of the applications involved in *Lilly* and *Fiers* and the present application. For example, the technique of polymerase chain reaction (PCR) was invented. Highly efficient cloning and DNA sequencing technology has been developed. Large databases of protein and nucleotide sequences have been compiled. Much of the raw material of the human and other genomes has been sequenced. With these remarkable advances one of skill in the art would recognize that, given the sequence information of SEQ ID NO:1, and the additional extensive detail provided by the subject application, the present inventors were in possession of the claimed polypeptide variants at the time of filing of this application.

4. Summary

The Office Action failed to base its written description inquiry "on whatever is now claimed." Consequently, the Action did not provide an appropriate analysis of the present claims and how they differ from those found not to satisfy the written description requirement in cases such as *Lilly* and *Fiers*. In particular, the claims of the subject application are fundamentally different from those found invalid in *Lilly* and *Fiers*. The subject matter of the present claims is defined in terms of the chemical structure of SEQ ID NO:1. The courts have stressed that structural features are important factors to consider in a written description analysis of claims to nucleic acids and proteins. In addition, the genus

of polypeptides defined by the present claims is adequately described, as evidenced by Brenner et al and consideration of the claims of the '740 patent involved in *Lilly*. Furthermore, there have been remarkable advances in the state of the art since the *Lilly* and *Fiers* cases, and these advances were given no consideration whatsoever in the position set forth by the Office Action. Therefore, withdrawal of the written description rejection under U.S.C. § 112, first paragraph is respectfully requested.

Enablement Rejections under 35 U.S.C. § 112, first paragraph

Claims 20 and 29 are rejected for allegedly failing to meet the requirements of 35 U.S.C. § 112, first paragraph, on the grounds that the specification does not provide an enabling disclosure commensurate in scope with the claims. In particular, the Examiner alleges that the specification "while being enabling for polypeptides comprising SEQ ID NO:1 or consisting of a fragment of SEQ ID NO:1 having endooliopeptidase activity, does not reasonably provide enablement for polypeptides consisting of any biologically active fragment of SEQ ID NO:1 " (Office Action page 9). Applicants respectfully traverse the rejection on the following grounds.

Claim 20 as currently pending recites biologically active fragments having endooligopeptidase activity. Given the sequence of SEQ ID NO:1, one of ordinary skill in the art could readily identify fragments of SEQ ID NO:1, using well known methods of sequence analysis, without any undue experimentation. The Examiner concedes that "recombinant techniques are known which could be used to make fragments of the polypeptide of SEQ ID NO:1 " (Office Action, page 10). Assays for endooligopeptidase activity were well known in the art at the time of filing of the instant application (For example, see the references of Hayashi et al. (1996) and Hayashi et al. (2000), cited by the Examiner, which describe assays for endooligopeptidase activity). Assays for protease activity are also described in the specification at page 51, line 26 through page 52, line 10. Thus, one of skill in the art could readily identify the claimed "biologically active fragments" without any undue experimentation." Therefore, withdrawal of the enablement rejection under U.S.C. § 112, first paragraph is respectfully requested.

Rejection under 35 U.S.C. § 102

Claims 20 and 29 are rejected under 35 U.S.C. § 102b as allegedly being anticipated by the references of Hayashi et al. (1996), Hayashi et al. (2000), and Sigma Catalog (1997). Claim 20 has been amended to recite variants that comprise "a naturally occurring amino acid sequence at least **95%** identical to an amino acid sequence of SEQ ID NO:1." Support for this amendment to the claim can be found in the specification at page 18, lines 15-18, which states:

> A preferred HPRAP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HPRAP amino acid sequence, and which contains at least one functional or structural characteristic of HPRAP.

Claim 20 has also been amended to remove the immunogenic fragment embodiment. Since the references do not disclose SEQ ID NO:1, nor a sequence 95% identical to SEQ ID NO:1, not every limitation of the claims is anticipated, and Applicants respectfully request withdrawal of the rejection under 35 U.S.C. § 102b.

CONCLUSION

In light of the above amendments and remarks, Applicants submit that the present application is fully in condition for allowance, and request that the Examiner withdraw the outstanding rejections. Early notice to that effect is earnestly solicited.

If the Examiner contemplates other action, or if a telephone conference would expedite allowance of the claims, Applicants invite the Examiner to contact Applicants' Attorney at (650) 855-0555.

Applicants believe that no fee is due with this communication. However, if the USPTO determines that a fee is due, the Commissioner is hereby authorized to charge Deposit Account No. **09-0108.**

Respectfully submitted,
INCYTE CORPORATION

Date: 27 June 2003

for Jenny Buchbinder
Reg. No. 48,588
Direct Dial Telephone: (650) 843-7212

Date: 27 June 2003

Cathleen M. Rocco
Reg. No. 46,172
Direct Dial Telephone: (650) 845-4587

Customer No.: 27904
3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 849-8886

VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE CLAIMS:

Claims 20, 31, 34, and 37 have been amended as follows:

20. (Once Amended) An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence of SEQ ID NO:1,

b) a polypeptide comprising a naturally occurring [an] amino acid sequence at least [70] 95% identical to an amino acid sequence of SEQ ID NO:1, wherein said polypeptide has endooligopeptidase activity, and

c) a biologically active fragment of a polypeptide having an amino acid sequence of SEQ ID NO:1, wherein said fragment has endooligopeptidase activity. [and

d) an immunogenic fragment of a polypeptide having an amino acid sequence of SEQ ID NO:1]

31. (Once Amended) A method for treating a disease or condition associated with decreased expression of functional [ABBR] HPRAP, comprising administering to a patient in need of such treatment the composition of claim 29.

34. (Once Amended) A method for treating a disease or condition associated with decreased expression of functional [HRAP] HPRAP, comprising administering to a patient in need of such treatment a composition of claim 33.

37. (Once Amended) A method for treating a disease or condition associated with overexpression of functional [HRAP] HPRAP, comprising administering to a patient in need of such treatment a composition of claim 36.

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6073–6078, May 1998
Biochemistry

Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships

STEVEN E. BRENNER*†‡, CYRUS CHOTHIA*, AND TIM J. P. HUBBARD§

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; and §Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, United Kingdom

Communicated by David R. Davies, National Institute of Diabetes, Bethesda, MD, March 16, 1998 (received for review November 12, 1997)

ABSTRACT **Pairwise sequence comparison methods have been assessed using proteins whose relationships are known reliably from their structures and functions, as described in the SCOP database [Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia C. (1995) *J. Mol. Biol.* 247, 536–540]. The evaluation tested the programs BLAST [Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* 215, 403–410], WU-BLAST2 [Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* 266, 460–480], FASTA [Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448], and SSEARCH [Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* 147, 195–197] and their scoring schemes. The error rate of all algorithms is greatly reduced by using statistical scores to evaluate matches rather than percentage identity or raw scores. The E-value statistical scores of SSEARCH and FASTA are reliable: the number of false positives found in our tests agrees well with the scores reported. However, the P-values reported by BLAST and WU-BLAST2 exaggerate significance by orders of magnitude. SSEARCH, FASTA ktup = 1, and WU-BLAST2 perform best, and they are capable of detecting almost all relationships between proteins whose sequence identities are >30%. For more distantly related proteins, they do much less well; only one-half of the relationships between proteins with 20–30% identity are found. Because many homologs have low sequence similarity, most distant relationships cannot be detected by any pairwise comparison method; however, those which are identified may be used with confidence.**

Sequence database searching plays a role in virtually every branch of molecular biology and is crucial for interpreting the sequences issuing forth from genome projects. Given the method's central role, it is surprising that overall and relative capabilities of different procedures are largely unknown. It is difficult to verify algorithms on sample data because this requires large data sets of proteins whose evolutionary relationships are known unambiguously and independently of the methods being evaluated. However, nearly all known homologs have been identified by sequence analysis (the method to be tested). Also, it is generally very difficult to know, in the absence of structural data, whether two proteins that lack clear sequence similarity are unrelated. This has meant that although previous evaluations have helped improve sequence comparison, they have suffered from insufficient, imperfectly characterized, or artificial test data. Assessment also has been problematic because high quality database sequence searching attempts to have both sensitivity (detection of homologs) and specificity (rejection of unrelated proteins); however, these complementary goals are linked such that increasing one causes the other to be reduced.

Sequence comparison methodologies have evolved rapidly, so no previously published tests has evaluated modern versions of programs commonly used. For example, parameters in BLAST (1) have changed, and WU-BLAST2 (2)—which produces gapped alignments—has become available. The latest version of FASTA (3) previously tested was 1.6, but the current release (version 3.0) provides fundamentally different results in the form of statistical scoring.

The previous reports also have left gaps in our knowledge. For example, there has been no published assessment of thresholds for scoring schemes more sophisticated than percentage identity. Thus, the widely discussed statistical scoring measures have never actually been evaluated on large databases of real proteins. Moreover, the different scoring schemes commonly in use have not been compared.

Beyond these issues, there is a more fundamental question: in an absolute sense, how well does pairwise sequence comparison work? That is, what fraction of homologous proteins can be detected using modern database searching methods?

In this work, we attempt to answer these questions and to overcome both of the fundamental difficulties that have hindered assessment of sequence comparison methodologies. First, we use the set of distant evolutionary relationships in the SCOP: Structural Classification of Proteins database (4), which is derived from structural and functional characteristics (5). The SCOP database provides a uniquely reliable set of homologs, which are known independently of sequence comparison. Second, we use an assessment method that jointly measures both sensitivity and specificity. This method allows straightforward comparison of different sequence searching procedures. Further, it can be used to aid interpretation of real database searches and thus provide optimal and reliable results.

Previous Assessments of Sequence Comparison. Several previous studies have examined the relative performance of different sequence comparison methods. The most encompassing analyses have been by Pearson (6, 7), who compared the three most commonly used programs. Of these, the Smith–Waterman algorithm (8) implemented in SSEARCH (3) is the oldest and slowest but the most rigorous. Modern heuristics have provided BLAST (1) the speed and convenience to make it the most popular program. Intermediate between these two is FASTA (3), which may be run in two modes offering either greater speed (ktup = 2) or greater effectiveness (ktup = 1). Pearson also considered different parameters for each of these programs.

To test the methods, Pearson selected two representative proteins from each of 67 protein superfamilies defined by the PIR database (9). Each was used as a query to search the database, and the matched proteins were marked as being homologous or unrelated according to their membership of PIR

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EPQ, errors per query.

†Present address: Department of Structural Biology, Stanford University, Fairchild Building D-109, Stanford, CA 94305-5126
‡To whom reprints requests should be addressed. e-mail: brenner@hyper.stanford.edu.

superfamilies. Pearson found that modern matrices and "lnscaling" of raw scores improve results considerably. He also reported that the rigorous Smith–Waterman algorithm worked slightly better than FASTA, which was in turn more effective than BLAST.

Very large scale analyses of matrices have been performed (10), and Henikoff and Henikoff (11) also evaluated the effectiveness of BLAST and FASTA. Their test with BLAST considered the ability to detect homologs above a predetermined score but had no penalty for methods which also reported large numbers of spurious matches. The Henikoffs searched the SWISS-PROT database (12) and used PROSITE (13) to define homologous families. Their results showed that the BLOSUM62 matrix (14) performed markedly better than the extrapolated PAM-series matrices (15), which previously had been popular.

A crucial aspect of any assessment is the data that are used to test the ability of the program to find homologs. But in Pearson's and the Henikoffs' evaluations of sequence comparison, the correct results were effectively unknown. This is because the superfamilies in PIR and PROSITE are principally created by using the same sequence comparison methods which are being evaluated. Interdependency of data and methods creates a "chicken and egg" problem, and means for example, that new methods would be penalized for correctly identifying homologs missed by older programs. For instance, immunoglobulin variable and constant domains are clearly homologous, but PIR places them in different superfamilies. The problem is widespread: each superfamily in PIR 48.00 with a structural homolog is itself homologous to an average of 1.6 other PIR superfamilies (16).

To surmount these sorts of difficulties, Sander and Schneider (17) used protein structures to evaluate sequence comparison. Rather than comparing different sequence comparison algorithms, their work focused on determining a length-dependent threshold of percentage identity, above which all proteins would be of similar structure. A result of this analysis was the HSSP equation; it states that proteins with 25% identity over 80 residues will have similar structures, whereas shorter alignments require higher identity. (Other studies also have used structures (18–20), but these focused on a small number of model proteins and were principally oriented toward evaluating alignment accuracy rather than homology detection.)

A general solution to the problem of scoring comes from statistical measures (i.e., E-values and P-values) based on the extreme value distribution (21). Extreme value scoring was implemented analytically in the BLAST program using the Karlin and Altschul statistics (22, 23) and empirical approaches have been recently added to FASTA and SSEARCH. In addition to being heralded as a reliable means of recognizing significantly similar proteins (24, 25), the mathematical tractability of statistical scores "is a crucial feature of the BLAST algorithm" (1). The validity of this scoring procedure has been tested analytically and empirically (see ref. 2 and references in ref. 24). However, all large empirical tests used random sequences that may lack the subtle structure found within biological sequences (26, 27) and obviously do not contain any real homologs. Thus, although many researchers have suggested that statistical scores be used to rank matches (24, 25, 28), there have been no large rigorous experiments on biological data to determine the degree to which such rankings are superior.

A Database for Testing Homology Detection. Since the discovery that the structures of hemoglobin and myoglobin are very similar though their sequences are not (29), it has been apparent that comparing structures is a more powerful (if less convenient) way to recognize distant evolutionary relationships than comparing sequences. If two proteins show a high degree of similarity in their structural details and function, it is very probable that they have an evolutionary relationship though their sequence similarity may be low.

The recent growth of protein structure information combined with the comprehensive evolutionary classification in the SCOP database (4, 5) have allowed us to overcome previous limitations. With these data, we can evaluate the performance of sequence comparison methods on real protein sequences whose relationships are known confidently. The SCOP database uses structural information to recognize distant homologs, the large majority of which can be determined unambiguously. These superfamilies, such as the globins or the immunoglobulins, would be recognized as related by the vast majority of the biological community despite the lack of high sequence similarity.

From SCOP, we extracted the sequences of domains of proteins in the Protein Data Bank (PDB) (30) and created two databases. One (PDB90D-B) has domains, which were all <90% identical to any other, whereas (PDB40D-B) had those <40% identical. The databases were created by first sorting all protein domains in SCOP by their quality and making a list. The highest quality domain was selected for inclusion in the database and removed from the list. Also removed from the list (and discarded) were all other domains above the threshold level of identity to the selected domain. This process was repeated until the list was empty. The PDB40D-B database contains 1,323 domains, which have 9,044 ordered pairs of distant relationships, or ~0.5% of the total 1,749,006 ordered pairs. In PDB90D-B, the 2,079 domains have 53,988 relationships, representing 1.2% of all pairs. Low complexity regions of sequence can achieve spurious high scores, so these were masked in both databases by processing with the SEG program (27) using recommended parameters: 12 1.8 2.0. The databases used in this paper are available from http://sss.stanford.edu/sss/, and databases derived from the current version of SCOP may be found at http://scop.mrc-lmb.cam.ac.uk/scop/.

Analyses from both databases were generally consistent, but PDB40D-B focuses on distantly related proteins and reduces the heavy overrepresentation in the PDB of a small number of families (31, 32), whereas PDB90D-B (with more sequences) improves evaluations of statistics. Except where noted otherwise, the distant homolog results here are from PDB40D-B. Although the precise numbers reported here are specific to the structural domain databases used, we expect the trends to be general.

Assessment Data and Procedure. Our assessment of sequence comparison may be divided into four different major categories of tests. First, using just a single sequence comparison algorithm at a time, we evaluated the effectiveness of different scoring schemes. Second, we assessed the reliability of scoring procedures, including an evaluation of the validity of statistical scoring. Third, we compared sequence comparison algorithms (using the optimal scoring scheme) to determine their relative performance. Fourth, we examined the distribution of homologs and considered the power of pairwise sequence comparison to recognize them. All of the analyses used the databases of structurally identified homologs and a new assessment criterion.

The analyses tested BLAST (1), version 1.4.9MP, and WU-BLAST2 (2), version 2.0a13MP. Also assessed was the FASTA package, version 3.0t76 (3), which provided FASTA and the SSEARCH implementation of Smith–Waterman (8). For SSEARCH and FASTA, we used BLOSUM45 with gap penalties −12/−1 (7, 16). The default parameters and matrix (BLOSUM62) were used for BLAST and WU-BLAST2.

The "Coverage Vs. Error" Plot. To test a particular protocol (comprising a program and scoring scheme), each sequence from the database was used as a query to search the database. This yielded ordered pairs of query and target sequences with associated scores, which were sorted, on the basis of their scores, from best to worst. The ideal method would have

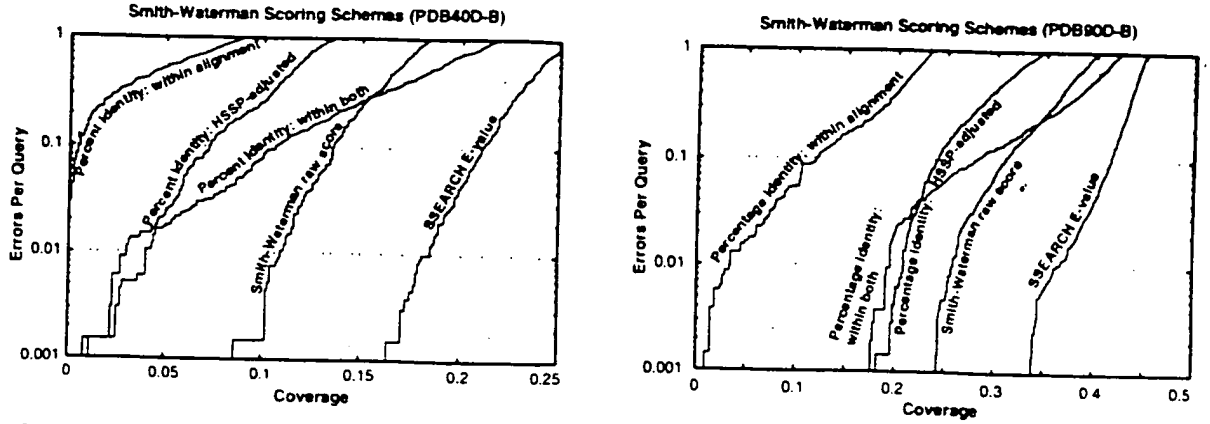


FIG. 1. Coverage vs. error plots of different scoring schemes for SSEARCH Smith–Waterman. (*A*) Analysis of PDB40D-B database. (*B*) Analysis of PDB90D-B database. All of the proteins in the database were compared with each other using the SSEARCH program. The results of this single set of comparisons were considered using five different scoring schemes and assessed. The graphs show the coverage and errors per query (EPQ) for statistical scores, raw scores, and three measures using percentage identity. In the coverage vs. error plot, the x axis indicates the fraction of all homologs in the database (known from structure) which have been detected. Precisely, it is the number of detected pairs of proteins with the same fold divided by the total number of pairs from a common superfamily. PDB40D-B contains a total of 9,044 homologs, so a score of 10% indicates identification of 904 relationships. The y axis reports the number of EPQ. Because there are 1,323 queries made in the PDB40D-B all-vs.-all comparison, 13 errors corresponds to 0.01, or 1% EPQ. The y axis is presented on a log scale to show results over the widely varying degrees of accuracy which may be desired. The scores that correspond to the levels of EPQ and coverage are shown in Fig. 4 and Table 1. The graph demonstrates the trade-off between sensitivity and selectivity. As more homologs are found (moving to the right), more errors are made (moving up). The ideal method would be in the lower right corner of the graph, which corresponds to identifying many evolutionary relationships without selecting unrelated proteins. Three measures of percentage identity are plotted. Percentage identity within alignment is the degree of identity within the aligned region of the proteins, without consideration of the alignment length. Percentage identity within both is the number of identical residues in the aligned region as a percentage of the average length of the query and target proteins. The HSSP equation (17) is $H = 290.15l^{-0.562}$ where l is length for $10 < l < 80$; $H > 100$ for $l < 10$; $H = 24.7$ for $l > 80$. The percentage identity HSSP-adjusted score is the percent identity within the alignment minus H. Smith–Waterman raw scores and E-values were taken directly from the sequence comparison program.

perfect separation, with all of the homologs at the top of the list and unrelated proteins below. In practice, perfect separation is impossible to achieve so instead one is interested in drawing a threshold above which there are the largest number of related pairs of sequences consistent with an acceptable error rate.

Our procedure involved measuring the coverage and error for every threshold. Coverage was defined as the fraction of structurally determined homologs that have scores above the selected threshold; this reflects the sensitivity of a method. Errors per query (EPQ), an indicator of selectivity, is the number of nonhomologous pairs above the threshold divided by the number of queries. Graphs of these data, called coverage vs. error plots, were devised to understand how protocols compare at different levels of accuracy. These graphs share effectively all of the beneficial features of Reciever Operating Characteristic (ROC) plots (33, 34) but better represent the high degrees of accuracy required in sequence comparison and the huge background of nonhomologs.

This assessment procedure is directly relevant to practical sequence database searching, for it provides precisely the information necessary to perform a reliable sequence database search. The EPQ measure places a premium on score consistency; that is, it requires scores to be comparable for different queries. Consistency is an aspect which has been largely

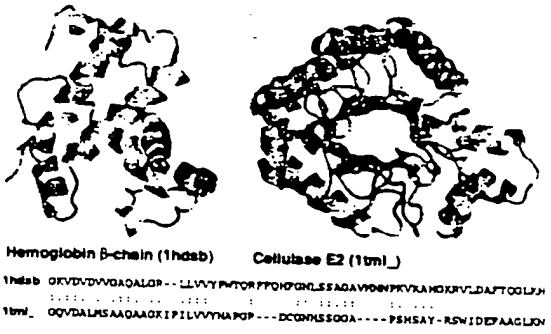


FIG. 2. Unrelated proteins with high percentage identity. Hemoglobin β-chain (PDB code 1hds chain b, ref. 38, *Left*) and cellulase E2 (PDB code 1tml, ref. 39, *Right*) have 39% identity over 64 residues, a level which is often believed to be indicative of homology. Despite this high degree of identity, their structures strongly suggest that these proteins are not related. Appropriately, neither the raw alignment score of 85 nor the E-value of 1.3 is significant. Proteins rendered by RASMOL (40).

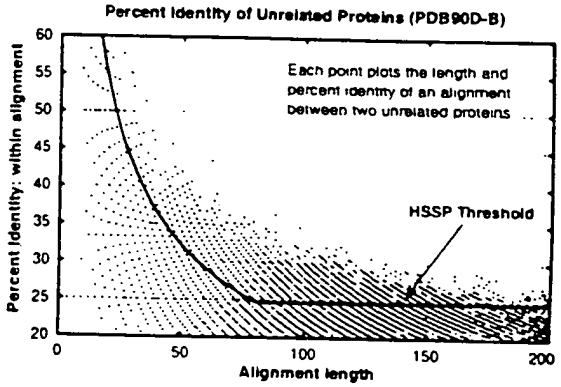


FIG. 3. Length and percentage identity of alignments of unrelated proteins in PDB90D-B: Each pair of nonhomologous proteins found with SSEARCH is plotted as a point whose position indicates the length and the percentage identity within the alignment. Because alignment length and percentage identity are quantized, many pairs of proteins may have exactly the same alignment length and percentage identity. The line shows the HSSP threshold (though it is intended to be applied with a different matrix and parameters).

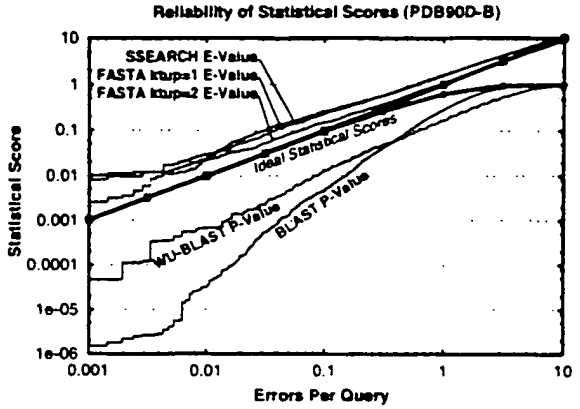


FIG. 4. Reliability of statistical scores in PDB90D-B: Each line shows the relationship between reported statistical score and actual error rate for a different program. E-values are reported for SSEARCH and FASTA, whereas P-values are shown for BLAST and WU-BLAST2. If the scoring were perfect, then the number of errors per query and the E-values would be the same, as indicated by the upper bold line. (P-values should be the same as EPQ for small numbers, and diverges at higher values, as indicated by the lower bold line.) E-values from SSEARCH and FASTA are shown to have good agreement with EPQ but underestimate the significance slightly. BLAST and WU-BLAST2 are overconfident, with the degree of exaggeration dependent upon the score. The results for PDB40D-B were similar to those for PDB90D-B despite the difference in number of homologs detected. This graph could be used to roughly calibrate the reliability of a given statistical score.

ignored in previous tests but is essential for the straightforward or automatic interpretation of sequence comparison results. Further, it provides a clear indication of the confidence that should be ascribed to each match. Indeed, the EPQ measure should approximate the expectation value reported by database searching programs, if the programs' estimates are accurate.

The Performance of Scoring Schemes. All of the programs tested could provide three fundamental types of scores. The first score is the percentage identity, which may be computed in several ways based on either the length of the alignment or the lengths of the sequences. The second is a "raw" or "Smith–Waterman" score, which is the measure optimized by the Smith–Waterman algorithm and is computed by summing the substitution matrix scores for each position in the alignment and subtracting gap penalties. In BLAST, a measure related to this score is scaled into bits. Third is a statistical score based on the extreme value distribution. These results are summarized in Fig. 1.

Sequence Identity. Though it has been long established that percentage identity is a poor measure (35), there is a common rule-of-thumb stating that 30% identity signifies homology. Moreover, publications have indicated that 25% identity can be used as a threshold (17, 36). We find that these thresholds, originally derived years ago, are not supported by present results. As databases have grown, so have the possibilities for chance alignments with high identity; thus, the reported cutoffs lead to frequent errors. Fig. 2 shows one of the many pairs of proteins with very different structures that nonetheless have high levels of identity over considerable aligned regions. Despite the high identity, the raw and the statistical scores for such incorrect matches are typically not significant. The principal reasons percentage identity does so poorly seem to be that it ignores information about gaps and about the conservative or radical nature of residue substitutions.

From the PDB90D-B analysis in Fig. 3, we learn that 30% identity is a reliable threshold for this database only for sequence alignments of at least 150 residues. Because one unrelated pair of proteins has 43.5% identity over 62 residues, it is probably necessary for alignments to be at least 70 residues in length before 40% is a reasonable threshold, for a database of this particular size and composition.

At a given reliability, scores based on percentage identity detect just a fraction of the distant homologs found by statistical scoring. If one measures the percentage identity in the aligned regions without consideration of alignment length, then a negligible number of distant homologs are detected. Use of the HSSP equation improves the value of percentage identity, but even this measure can find only 4% of all known homologs at 1% EPQ. In short, percentage identity discards most of the information measured in a sequence comparison.

Raw Scores. Smith–Waterman raw scores perform better than percentage identity (Fig. 1), but ln-scaling (7) provided no notable benefit in our analysis. It is necessary to be very precise when using either raw or bit scores because a 20% change in cutoff score could yield a tenfold difference in EPQ. However, it is difficult to choose appropriate thresholds because the reliability of a bit score depends on the lengths of the proteins matched and the size of the database. Raw score thresholds also are affected by matrix and gap parameters.

Statistical Scores. Statistical scores were introduced partly to overcome the problems that arise from raw scores. This scoring scheme provides the best discrimination between homologous proteins and those which are unrelated. Most

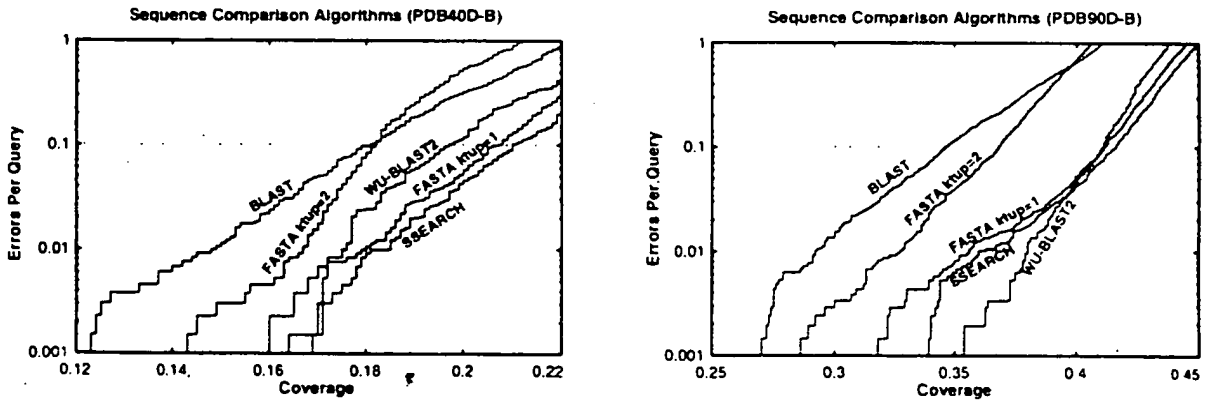


FIG. 5. Coverage vs. error plots of different sequence comparison methods: Five different sequence comparison methods are evaluated, each using statistical scores (E- or P-values). (*A*) PDB40D-B database. In this analysis, the best method is the slow SSEARCH, which finds 18% of relationships at 1% EPQ. FASTA ktup = 1 and WU-BLAST2 are almost as good. (*B*) PDB90D-B database. The quick WU-BLAST2 program provides the best coverage at 1% EPQ on this database, although at higher levels of error it becomes slightly worse than FASTA ktup = 1 and SSEARCH.

likely, its power can be attributed to its incorporation of more information than any other measure; it takes account of the full substitution and gap data (like raw scores) but also has details about the sequence lengths and composition and is scaled appropriately.

We find that statistical scores are not only powerful, but also easy to interpret. SSEARCH and FASTA show close agreement between statistical scores and actual number of errors per query (Fig. 4). The expectation value score gives a good, slightly conservative estimate of the chances of the two sequences being found at random in a given query. Thus, an E-value of 0.01 indicates that roughly one pair of nonhomologs of this similarity should be found in every 100 different queries. Neither raw scores nor percentage identity can be interpreted in this way, and these results validate the suitability of the extreme value distribution for describing the scores from a database search.

The P-values from BLAST also should be directly interpretable but were found to overstate significance by more than two orders of magnitude for 1% EPQ for this database. Nonetheless, these results strongly suggest that the analytic theory is fundamentally appropriate. WU-BLAST2 scores were more reliable than those from BLAST, but also exaggerate expected confidence by more than an order of magnitude at 1% EPQ.

Overall Detection of Homologs and Comparison of Algorithms. The results in Fig. 5*A* and Table 1 show that pairwise sequence comparison is capable of identifying only a small fraction of the homologous pairs of sequences in PDB40D-B. Even SSEARCH with E-values, the best protocol tested, could find only 18% of all relationships at a 1% EPQ. BLAST, which identifies 15%, was the worst performer, whereas FASTA ktup = 1 is nearly as effective as SSEARCH. FASTA ktup = 2 and WU-BLAST2 are intermediate in their ability to detect homologs. Comparison of different algorithms indicates that those capable of identifying more homologs are generally slower. SSEARCH is 25 times slower than BLAST and 6.5 times slower than FASTA ktup = 1. WU-BLAST2 is slightly faster than FASTA ktup = 2, but the latter has more interpretable scores.

In PDB90D-B, where there are many close relationships, the best method can identify only 38% of structurally known homologs (Fig. 5*B*). The method which finds that many relationships is WU-BLAST2. Consequently, we infer that the differences between FASTA kup = 1, SSEARCH, and WU-BLAST2 programs are unlikely to be significant when compared with variation in database composition and scoring reliability.

Fig. 6 helps to explain why most distant homologs cannot be found by sequence comparison: a great many such relationships have no more sequence identity than would be expected by chance. SSEARCH with E-values can recognize >90% of the homologous pairs with 30–40% identity. In this region, there are 30 pairs of homologous proteins that do not have significant E-values, but 26 of these involve sequences with <50 residues. Of sequences having 25–30% identity, 75% are identified by SSEARCH E-values. However, although the number of homologs grows at lower levels of identity, the detection falls off sharply: only 40% of homologs with 20–25% identity

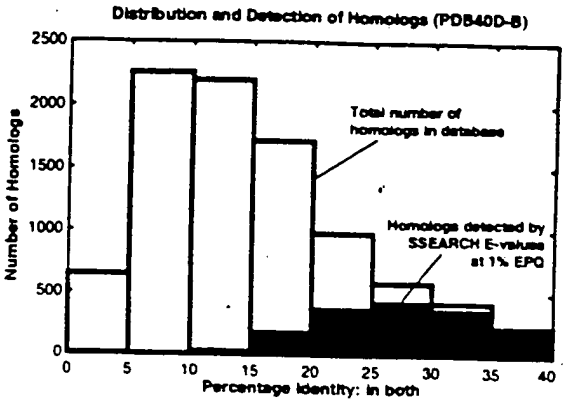


FIG. 6. Distribution and detection of homologs in PDB40D-B. Bars show the distribution of homologous pairs PDB40D-B according to their identity (using the measure of identity in both). Filled regions indicate the number of these pairs found by the best database searching method (SSEARCH with E-values) at 1% EPQ. The PDB40D-B database contains proteins with <40% identity, and as shown on this graph, most structurally identified homologs in the database have diverged extremely far in sequence and have <20% identity. Note that the alignments may be inaccurate, especially at low levels of identity. Filled regions show that SSEARCH can identify most relationships that have 25% or more identity, but its detection wanes sharply below 25%. Consequently, the great sequence divergence of most structurally identified evolutionary relationships effectively defeats the ability of pariwise sequence comparison to detect them.

are detected and only 10% of those with 15–20% can be found. These results show that statistical scores can find related proteins whose identity is remarkably low; however, the power of the method is restricted by the great divergence of many protein sequences.

After completion of this work, a new version of pairwise BLAST was released: BLASTGP (37). It supports gapped alignments, like WU-BLAST2, and dispenses with sum statistics. Our initial tests on BLASTGP using default parameters show that its E-values are reliable and that its overall detection of homologs was substantially better than that of ungapped BLAST, but not quite equal to that of WU-BLAST2.

CONCLUSION

The general consensus amongst experts (see refs. 7, 24, 25, 27 and references therein) suggests that the most effective sequence searches are made by (*i*) using a large current database in which the protein sequences have been complexity masked and (*ii*) using statistical scores to interpret the results. Our experiments fully support this view.

Our results also suggest two further points. First, the E-values reported by FASTA and SSEARCH give fairly accurate estimates of the significance of each match, but the P-values provided by BLAST and WU-BLAST2 underestimate the true

Table 1. Summary of sequence comparison methods with PDB40D-B

Method	Relative Time*	1% EPQ Cutoff	Coverage at 1% EPQ
SSEARCH % identity: within alignment	25.5	>70%	<0.1
SSEARCH % identity: within both	25.5	34%	3.0
SSEARCH % identity: HSSP-scaled	25.5	35% (HSSP + 9.8)	4.0
SSEARCH Smith–Waterman raw scores	25.5	142	10.5
SSEARCH E-values	25.5	0.03	18.4
FASTA ktup = 1 E-values	3.9	0.03	17.9
FASTA ktup = 2 E-values	1.4	0.03	16.7
WU-BLAST2 P-values	1.1	0.003	17.5
BLAST P-values	1.0	0.00016	14.8

*Times are from large database searches with genome proteins.

extent of errors. Second, SSEARCH, WU-BLAST2, and FASTA ktup = 1 perform best, though BLAST and FASTA ktup = 2 detect most of the relationships found by the best procedures and are appropriate for rapid initial searches.

The homologous proteins that are found by sequence comparison can be distinguished with high reliability from the huge number of unrelated pairs. However, even the best database searching procedures tested fail to find the large majority of distant evolutionary relationships at an acceptable error rate. Thus, if the procedures assessed here fail to find a reliable match, it does not imply that the sequence is unique; rather, it indicates that any relatives it might have are distant ones.**

**Additional and updated information about this work, including supplementary figures, may be found at http://sss.stanford.edu/sss/.

The authors are grateful to Drs. A. G. Murzin, M. Levitt, S. R. Eddy, and G. Mitchison for valuable discussion. S.E.B. was principally supported by a St. John's College (Cambridge, UK) Benefactors' Scholarship and by the American Friends of Cambridge University. S.E.B. dedicates his contribution to the memory of Rabbi Albert T. and Clara S. Bilgray.

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
2. Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* **266**, 460–480.
3. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
4. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* **247**, 536–540.
5. Brenner, S. E., Chothia, C., Hubbard, T. J. P. & Murzin, A. G. (1996) *Methods Enzymol.* **266**, 635–643.
6. Pearson, W. R. (1991) *Genomics* **11**, 635–650.
7. Pearson, W. R. (1995) *Protein Sci.* **4**, 1145–1160.
8. Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* **147**, 195–197.
9. George, D. G., Hunt, L. T. & Barker, W. C. (1996) *Methods Enzymol.* **266**, 41–59.
10. Vogt, G., Etzold, T. & Argos, P. (1995) *J. Mol. Biol.* **249**, 816–831.
11. Henikoff, S. & Henikoff, J. G. (1993) *Proteins* **17**, 49–61.
12. Bairoch, A. & Apweiler, R. (1996) *Nucleic Acids Res.* **24**, 21–25.
13. Bairoch, A., Bucher, P. & Hofmann, K. (1996) *Nucleic Acids Res.* **24**, 189–196.
14. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
15. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (National Biomedical Research Foundation, Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
16. Brenner, S. E. (1996) Ph.D. thesis. (University of Cambridge, UK).
17. Sander, C. & Schneider, R. (1991) *Proteins* **9**, 56–68.
18. Johnson, M. S. & Overington, J. P. (1993) *J. Mol. Biol.* **233**, 716–738.
19. Barton, G. J. & Sternberg, M. J. E. (1987) *Protein Eng.* **1**, 89–94.
20. Lesk, A. M., Levitt, M. & Chothia, C. (1986) *Protein Eng.* **1**, 77–78.
21. Arratia, R., Gordon, L. & M. W. (1986) *Ann. Stat.* **14**, 971–993.
22. Karlin, S. & Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2264–2268.
23. Karlin, S. & Altschul, S. F. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5873–5877.
24. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6**, 119–129.
25. Pearson, W. R. (1996) *Methods Enzymol.* **266**, 227–258.
26. Lipman, D. J., Wilbur, W. J., Smith, T. F. & Waterman, M. S. (1984) *Nucleic Acids Res.* **12**, 215–226.
27. Wootton, J. C. & Federhen, S. (1996) *Methods Enzymol.* **266**, 554–571.
28. Waterman, M. S. & Vingron, M. (1994) *Stat. Science* **9**, 367–381.
29. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13**, 669–678.
30. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987) in *Crystallographic Databases: Information Content, Software Systems, Scientific Applications*, eds. Allen, F. H., Bergerhoff, G. & Sievers, R. (Data Comm. Intl. Union Crystallogr., Cambridge, UK), pp. 107–132.
31. Brenner, S. E., Chothia, C. & Hubbard, T. J. P. (1997) *Curr. Opin. Struct. Biol.* **7**, 369–376.
32. Orengo, C., Michie, A., Jones S. Jones D. T. Swindells M. B. & Thornton, J. (1997) *Structure (London)* **5**, 1093–1108.
33. Zweig, M. H. & Campbell, G. (1993) *Clin. Chem.* **39**, 561–577.
34. Gribskov, M. & Robinson, N. L. (1996) *Comput. Chem.* **20**, 25–33.
35. Fitch, W. M. (1966) *J. Mol. Biol.* **16**, 9–16.
36. Chung, S. Y. & Subbiah, S. (1996) *Structure (London)* **4**, 1123–1127.
37. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
38. Girling, R., Schmidt, W., Jr. Houston, T., Amma, E. & Huisman, T. (1979) *J. Mol. Biol.* **131**, 417–433.
39. Spezio, M., Wilson, D. & Karplus, P. (1993) *Biochemistry* **32**, 9906–9916
40. Sayle, R. A. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20**, 374–376.

Biol. Chem. Hoppe-Seyler, Vol. 377, pp. [illegible]–291, May 1996 · Copyright © by Walter de Gruyte · Berlin · New York

Species Specificity of Thimet Oligopeptidase (EC 3.4.24.15)

Mirian A. F. Hayashi[1], Marcelo D. Gomes[1], Nancy A. Rebouças[2], Beatriz L. Fernandes[3], Emer S. Ferro[2] and Antonio C.M. de Camargo[1,*]

[1] Laboratory of Biochemistry and Biophysics of the Institute Butantan, 05503-900, São Paulo, Brazil
[2] Department of Physiology and
[3] Department of Microbiology,
Institute of Biomedical Sciences, University of São Paulo, 05508, São Paulo, Brazil

* Corresponding author

The recombinant rat testes metallo-endooligopeptidase (EC 3.4.24.15) and the rabbit brain endooligopeptidase A (formerly EC 3.4.22.19) were compared, side-by-side, in view of their striking similarities in both the physicochemical features and the specificities for oligopeptides. Concerning the tissue distribution in rat and rabbit, no relation between the levels of enzyme activity in cytosol and the levels of metallo-endooligopeptidase 24.15 mRNA could be established. The results suggest that the predominant neuropeptide-metabolizing activity attributed to the metallo-endooligopeptidase 24.15 is performed by, at least, two distinct cytosolic enzymes, one predominant in rat testes and the other in rabbit brain and testes, and possibly also in rat brain. Both enzymes are activated by dithiothreitol and irreversibly inhibited by a SH-affinity labeling dynorphin-related compound, but they are not inhibited by EDTA in a concentration dependent manner. Both enzymes exhibit the same specificity toward several bioactive peptides, except for LH-RH and substance P, which are only hydrolysed by the rat testes enzyme. Taken together, these results lead us to conclude that it is unlikely that the recombinant rat testes metallo-endooligopeptidase 24.15 and the rabbit brain endooligopeptidase A are the same molecule although they might belong to the same family of oligopeptidases.

Key words: Brain peptidases / Neuropeptide metabolism / Thimet oligopeptidase.

Introduction

During the past four decades an increasing number of bioactive oligopeptides has been discovered and was included in the class of the intercellular signaling molecules which regulate several neural, endocrine and immune processes. They are hormones, neurotransmitters, neuro- or immunomodulators which are all generated by partial proteolysis from protein precursors, giving rise to biologically active peptides (Hardie, 1992). The modulation of the action of these bioactive peptides depends on the activity of tissue peptidases which can either inactivate or convert the peptide into a smaller bioactive product.

In the early 70's two thiol-activated endopeptidases were isolated from cytosol of rabbit brain using the nonapeptide bradykinin (Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg) as substrate (Camargo *et al.*, 1973; Oliveira *et al.*, 1976). These enzymes, displaying similar physicochemical properties but exhibiting distinct preferences for specific peptide bonds of bradykinin (Oliveira *et al.*, 1976), were further purified to apparent homogeneity (Carvalho and Camargo, 1981), representing the first enzymes showing selectivity for oligopeptides (Camargo *et al.*, 1979a; Barrett and Rawlings, 1992). They were generically named endo-oligopeptidases A (EOPA) [originally classified as a cysteine-protease EC 3.4.22.19], and endo-oligopeptidase B or proline-endopeptidase (EC 3.4.21.26). These enzymes can be distinguished from typical proteinases not only because of their strict selectivity for oligopeptides, but also because the sub-site interactions between the substrate and the enzyme do not seem to play a dominant role in the enzyme specificity (Camargo *et al.*, 1987; Camargo *et al.*, 1994). In 1992, the International Union of Biochemistry and Molecular Biology recognized the oligopeptidases as a distinct family of proteolytic enzymes displaying selectivity towards oligopeptides (Enzyme Nomenclature Recommendations, 1992).

The rabbit brain EOPA, a 70 kDa, EDTA-insensitive cysteine-peptidase (Shaw, 1990) hydrolyzes selectively peptides of 7 to 13 amino acid residues (Camargo *et al.*, 1987; Camargo *et al.*, 1994). This enzyme specifically hydrolyzes the Phe^5-Ser^6 bond of bradykinin (Camargo *et al.*, 1973) and the Arg^8-Arg^9 bond of neurotensin (Camargo *et al.*, 1983), and releases [Met^5] or [Leu^5] enkephalin from a number of enkephalin-containing peptides (Camargo *et al.*, 1985; Camargo *et al.*, 1987).

In 1983, Orlowski and his collaborators described the purification of a similar endopeptidase from cytosol of rat brain using a chromogenic substrate, Bz-Gly-Ala-Ala-Phe-pAB (Orlowski *et al.*, 1983). The most significant difference between the rat and the rabbit endo-oligopeptidases was the sensitivity of the rat enzyme to chelating compounds, thus justifying its classification among the metallo-peptidases (EC 3.4.24.15), whereas the endo-oligopeptidase A had received another classification (EC 3.4.22.19) among the thiol-proteases (Enzyme Nomenclature Committee, 1989). Except for this difference, which was attributed to enzymes obtained from distinct

animal species (Orlowski *et al.*, 1983), the rabbit brain EOPA and the rat metallo-endooligopeptidase (EOP 24.15) were strikingly similar both in their physicochemical properties (Carvalho and Camargo, 1981; Orlowski *et al.*, 1983) and in their specificities towards a number of bioactive peptides (Camargo *et al.*, 1973; Camargo *et al.*, 1985; Camargo *et al.*, 1987; Orlowski *et al.*, 1983; Chu and Orlowski, 1985). In the early 90's Orlowski and his group succeeded in obtaining the cDNA coding for EOP 24.15 from a rat testes cDNA library (Pierotti *et al.*, 1990). The primary sequence deduced from the cloned rat testes EOP 24.15 cDNA revealed the presence of the metallo-endopeptidases motif HEXXH (Pierotti *et al.*, 1990). In 1992, Barrett and co-workers postulated that the rabbit brain EOPA and the EOP 24.15 were the same enzyme (Barrett and Rawlings, 1992). A single name, thimet-oligopeptidase (thiol-activated metallo-endopeptidase), was suggested to include both the brain EOPA and the rat testes EOP 24.15.

The following dissimilarities between the EOPA and the EOP 24.15 led us to investigate whether they could be due to enzymes obtained from distinct animal sources (as suggested by Orlowski *et al.*, 1983):

1) The EOP 24.15 and not the EOPA is inhibited by EDTA (Orlowski *et al.*, 1983; Camargo *et al.*, 1973);

2) Pierotti *et al.* (1990) showed that there is low concentration of rat testes EOP 24.15 mRNA in rat brain;

3) the immunostaining of EOP 24.15 indicated the presence of this enzyme in the nucleus but not in the cytosol of rat brain (Healy and Orlowski, 1992) where more than 70% of the enzyme activity is located (Oliveira *et al.*, 1990);

4) the anti-[rat brain EOPA] antiserum was able to precipitate the activity of EOPA but not the activity of EOP 24.15 from the cytosol of rat brain (Toffoletto *et al.*, 1988);

5) the EOP 24.15 was referred to as being able to hydrolyze the LH-RH, substance P and angiotensin II (Orlowski *et al.*, 1983; Chu and Orlowski, 1985; Lasdun *et al.*, 1989), none of which are substrates for EOPA (Camargo *et al.*, 1979b; Camargo *et al.*, 1982; Camargo *et al.*, 1983).

In this paper we evaluated, side-by-side, a number of properties of the recombinant rat testes EOP 24.15 and the purified rabbit brain EOPA, and we concluded that it is unlikely that they are the same molecule, although they could be members of the oligopeptidase family (Barrett *et al.*, 1995).

Results

Enzyme Activity in Brain and Testes Cytosol of Rat and Rabbit

The EOP 24.15 activities determined by fluorimetric assay in the cytosol of rat and rabbit brain (0.71 and 0.86 mU/mg, respectively), were about three times lower than those found in the cytosol of rat and rabbit testes (2.4 and 2.1 mU/mg, respectively).

Northern Blot Analysis

Hybridization of the recombinant rat testes EOP 24.15 radioactive probe with total RNA from rat brain and testes showed a single band of approximately 3.1 kb, which was 23 times stronger in testes (Figure 1A, lane 1) than in brain (Figure 1A, lane 2). However, when total RNA from rabbit brain and testes were hybridized to the same probe, no signal could be detected (Figure 1A, lanes 3 and 4). Nevertheless, when purified mRNA from rabbit brain and testes were used, a single band corresponding to 3.1 kb was detected in testes (Figure 1B, lane 5), while a weak signal was detected in brain after a 3 days exposure (Figure 1B, lane 6). The integrity of rat and rabbit brain and testes RNA was confirmed by hybridization of a probe specific for rabbit β-actin to the total RNA and purified mRNA preparations used (Figures 1C and D, lanes 1 to 6).

Western Blot Analysis

Using anti-[recombinant EOP 24.15] antiserum, the comparative Western blot analysis (Figure 2) of 2 mU of each a cytosolic fraction of rat testes (lane 1), the purified recombinant EP 24.15 (lane 2), the purified rabbit brain EOPA (lane 3), and a cytosolic fraction of rat brain (lane 4), showed the presence of the 77.5 kDa band only in the

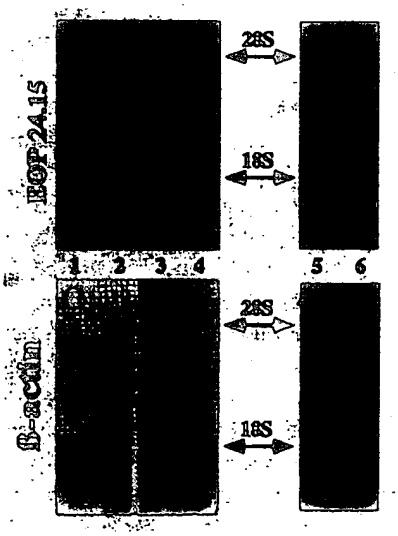


Fig. 1 Northern Blot Analysis of Recombinant Rat Testes EOP 24.15.

Total RNAs (10 μg) from rat and rabbit brain and testes, and purified mRNA (5 μg) from rabbit brain and testes were fractionated by electrophoresis, blotted onto nylon membrane and hybridized to the ^{32}P-labeled recombinant rat testes EOP 24.15 cDNA (Panels A and B), and to a ^{32}P-labeled 319 bp PCR-fragment of β-actin cDNA (Panels C and D). Panel A, total RNA: lane 1, rat testes; lane 2, rat brain; lane 3, rabbit testes; lane 4, rabbit brain. Panel B, mRNA: lane 5, rabbit testes; lane 6, rabbit brain. Panel C, total RNA: lane 1, rat testes; lane 2, rat brain; lane 3, rabbit testes; lane 4, rabbit brain. Panel D, mRNA: lane 5, rabbit testes; lane 6, rabbit brain. The migration of ribosomal RNAs (18S and 28S) is indicated. Autoradiographic exposures were for 24 hours except for the blot of Panel B, which was exposed for 72 hours.

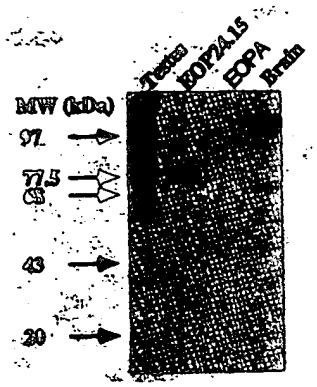


Fig. 2 Western Blot Analysis of the Cytosols of Rat Brain and Rat Testes, Purified Rabbit Brain EOPA and Purified Recombinant Rat Testes EOP 24.15.
Samples of 2 mU of each, cytosol of rat testes (0.15 mg) (lane 1), recombinant rat testes EOP 24.15 (0.27 μg) (lane 2), purified rabbit brain EOPA (0.21 μg) (lane 3), cytosol of rat brain (0.6 mg) (lane 4), and M_r standards were subjected to SDS-PAGE, and electrotransferred to a nitrocellulose membrane for 18 hours at 180 mA using a Trans-Blot cell. The nitrocellulose membranes were subsequently soaked in blot buffer for 4 hours before incubation for 1 hour with 1:500 dilution of anti-[recombinant rat testes EOP 24.15] antiserum. The Western blot was developed using an alkaline phosphatase conjugated antimouse IgG diluted 1:7500 in 3% BSA for 30 min and reacted with BCIP/NBT. Control experiments were conducted using Balb-c mouse pre-immune antiserum.

cytosol of rat testes but not in the cytosol of rat brain. The purified rabbit brain EOPA was not recognized by this antiserum. The presence of two other bands of 69 and 112 kDa were also present in the Western blot of cytosol of rat brain.

Eff ct of EDTA and DTT

Concentrations of 1 to 20 mM of EDTA in the incubation medium did not significantly affect the enzyme activity of either rabbit brain EOPA nor recombinant rat testes EOP 24.15 (Figure 3, left panel). On the other hand, both enzymes were activated to the same extent by dithiothreitol (DTT) (Figure 3, right panel), the highest activation occurring at the concentration of 0.5 mM (to a level about 90% above the control). This thiol compound was inhibitory for both enzymes at concentrations higher than 1 mM (data not shown).

Inhibition of the Recombinant Rat Testes EOP 24.15 Activity by YGGFLRRC(Npys)R(NH)$_2$ and by Cys(Npys)

The dynorphin Cys(Npys)-derived peptide but not the Cys(Npys) compound produced a rapid and irreversible inhibition of EOP 24.15. Figure 4 shows that YGGFLRRC(Npys)R(NH)$_2$ produced 95% inhibition of the recombinant EOP 24.15 after 1 min pre-incubation. This inhibition was partially reversed (85%) by the addition of 0.5 mM DTT. Figure 4 also shows that Cys(Npys) only produced 15% inhibition after 60 min of pre-incubation.

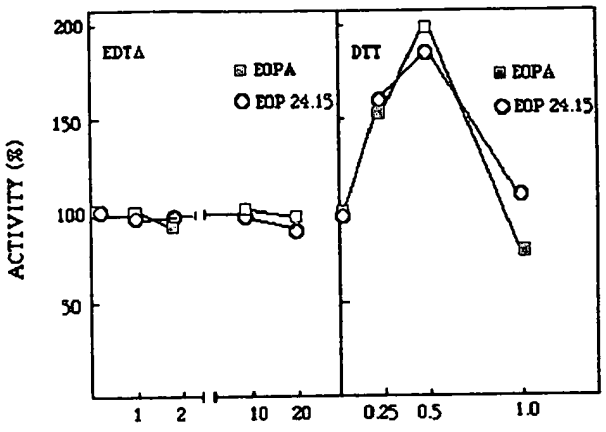


Fig. 3 Effect of EDTA and DTT on Recombinant Rat Testes EOP 24.15 and Rabbit Brain EOPA.
Approximately 1 mU of either purified rabbit brain EOPA or recombinant rat testes EOP 24.15 were pre-incubated with EDTA or DTT at 22 °C for 10 minutes in 200 μl of 20 mM Tris-HCl buffer, pH 7.5. After pre-incubation, 20 nmol of bradykinin were introduced to the medium and incubated at 37 °C for 10 min. The reaction was stopped with 10 μl TFA and the concentration of the remaining bradykinin was determined by HPLC (see Methods). The results represent the mean of the values of three independent experiments.

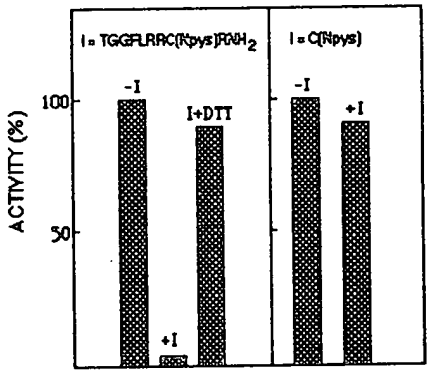


Fig. 4 Inhibition of the Recombinant Rat Testes EOP 24.15 Activity by YGGFLRRC(Npys)R(NH)$_2$ and by Cys(Npys).
Purified recombinant rat testes EOP 24.15 (approximately 2 mU) was pre-incubated for 1 min at 37 °C with 10 mM of the inhibitors in 100 μl of 20 mM Tris-HCl buffer, pH 7.5. Aliquots of 10 μl were diluted in 500 μl of same buffer containing 16 μM of QF-ERP$_7$ and assayed fluorimetrically at 37 °C (Gomes *et al.*, 1993). The effect of DTT was performed by addition of 0.5 mM of DTT to the enzyme which had been pre-incubated for 1 min with YGGFLRRC-(Npys)-R(NH)$_2$. The results represent the mean of the values of three independent experiments.

Specificities of the Recombinant EOP 24.15 and EOPA

The peptides bradykinin, neurotensin and the opioid peptide metorphinamide (Table 1) were cleaved by both recombinant rat testes EOP 24.15 and rabbit brain EOPA with approximately the same rates and at the same peptide bonds. The peptide angiotensin II was resistant to

Table 1 Relative Rate of Hydrolysis and Sites of Peptide Bond Cleavages by Recombinant Rat Testes EOP 24.15 and by Rabbit Brain EOPA.

Peptides	Cleavage	Hydrolysis* (%)	
		EOP 24.15	EOPA
Bradykinin	RPPGF ↑ ↑ SPFR	100	90
24.15 Substrate	Bz-G ↑ ↑ AAF-pAB	12	3
Neurotensin	<ELYEDKPR ↑ ↑ RPYIL	85	80
Metorphinamide	YGGFM ↑ ↑ RRV-(NH2)	115	130
Substance P	RPKPQQF ↑ F ↑ GLM-(NH_2)	11	0
LH-RH	<EHWSY ↑ GLRPG-(NH_2)	16	0
Angiotensin II	DRVYIHPF	0	0
Suc. bis bradykinin	CO-RPPGFSPFR | $(CH2)_2$ | CO-RPPGFSPFR	0	0
Dynorphin A_{1-17}	YGGFLRRIRPKLKWNNQ	0	0
BAM-22P	YGGFMRRVGRPEWWMDYQKRYG	0	0
b-endorphin	YGGFMTSEKSQTPLVTLFKNAIIKNAHKKGQ	0	0

* 100% represent the hydrolysis of 15 nmol of bradykinin / 12 min at 37 °C in 20 mM Tris-HCl, pH 7.5, containing 0.5 mM DTT and 35 nmol of substrate; ↑ Hydrolyzed only by EOP 24.15; ↑↑ Hydrolyzed by both enzymes.

hydrolysis by both enzymes. The recombinant rat testes EOP 24.15 was able to hydrolyze LH-RH at the Tyr-Gly bond and substance P either at the Phe-Phe or Phe-Gly bonds. Rabbit brain EOPA did not hydrolyze LH-RH or substance P even after a prolonged incubation time (12 hours). In addition, both enzymes were unable to hydrolyze peptides larger than neurotensin (13 amino acid residues) such as BAM-22P, β-endorphin, dynorphin A_{1-17} or the suc.bis-bradykinin.

Discussion

Exhaustive efforts have been made in our laboratory to isolate a homogeneous preparation of rabbit brain EOPA by conventional methods. Regardless of the fact that the nervous tissue is the most obvious tissue to search for neuropeptide-metabolizing peptidases, the brain EOPA has not yet been fully characterized. The main reason is that the purified brain EOPA contains a significant proportion of inactive proteins with the same molecular mass and isoelectric point. Similar difficulties were reported by Pierotti *et al.* (1990) for the isolation of rat brain EOP 24.15. These reasons led Orlowski and his collaborators to exclude the brain as an enzyme source, and to isolate the enzyme from rat testes, a tissue rich in EOP 24.15 activity displaying similar properties as the brain enzyme (Orlowski *et al.*, 1989). It was assumed that the brain EOPA-like activity was performed by the brain EOP 24.15, thus implying that both the EOPA and the EOP 24.15 were the same enzyme (Barrett and Rawlings, 1992). However, the results presented here and elsewhere indicate that an enzyme different from EOP 24.15 (most likely the EOPA) may be the predominant enzyme displaying the EOPA-like activity in the cytosol of rat or rabbit brain. The evidences are the following: the EOPA-like activity in rat brain is about 3-fold lower than the activity in rat testes whereas the EOP 24.15 mRNA levels were more than 20-fold lower in rat brain. This discrepancy was also reported by Pierotti *et al.* (1990). The lack of correspondence between the EOPA-like activity and the level of EOP 24.15 mRNA was even more pronounced when tissues of rat and rabbit were compared, thus suggesting that the distribution of EOP 24.15 is species-specific: no hybridization signal was observed in total RNA preparations of rabbit testes and brain, whereas a weak signal was obtained after a 20-fold enrichment of rabbit brain mRNA. In addition, the level of EOP 24.15 mRNA was significantly lower in rabbit testes as compared to rat testes, regardless of the fact that the EOPA-like activity of the cytosol of testes of both animals was approximately the same.

Further support to the suggestion that the predominant enzyme responsible for EOPA-like activity in the nervous tissue is different from the EOP 24.15 was brought by Western blot analysis using monospecific antiserum against recombinant rat testes EOP 24.15. In contrast to the high intensity of the 77 kDa protein band found when 2 mU of EOPA-like activity from cytosol of rat testes were analyzed, no visible 77 kDa band could be detected using the same amount of enzyme activity from cytosol of either rat or rabbit brain. The two other bands, one larger and the other smaller than the EOP 24.15, detected in rat brain cytosol are likely to correspond to EOP 24.15-homologous proteins (for review see Barrett *et al.*, 1995).

Furthermore, previous reports involving immunochemical detection of EOP 24.15, support the suggestion that

the EOP 24.15 is not the predominant enzyme responsible for the EOPA-like activity in cytosol of rat brain:

(i) The immunostaining of EOP 24.15 in nervous tissue had demonstrated that this enzyme is confined to the nucleus and not to the cytosol of the neurones (Healy and Orlowski, 1992) where more than 70% of the EOPA activity is located (Oliveira *et al.*, 1990);

(ii) the antibody raised against the rat brain EOPA was able to immunoprecipitate the EOPA but not the EOP 24.15 activity from the cytosol of rat brain (Toffoletto *et al.*, 1988).

It is well known that, under physiological conditions, the mRNA levels and the correspondent expression levels are often found in good correlation (Shields and Wickramasinghe, 1995; Piechaczyk *et al.*, 1984; Dauch *et al.*, 1995). Assuming that this correlation can also be applied to the expression of EOP 24.15 then other peptidases besides the EOP 24.15 would be necessary to account for the overall enzyme activity detected not only in rat brain but also in the testes and brain of rabbit. Moreover, the existence of isoforms of many proteins is well documented: the ion channel proteins (Orlowski *et al.*, 1992; Collins *et al.*, 1993) and the receptors proteins (Nakagawa *et al.*, 1991; Park *et al.*, 1994) are expressed differentially in various tissues and animals, probably contributing to their specific functions.

The comparison of the other properties of the recombinant rat testes EOP 24.15 and the rabbit brain EOPA indicated that there are more similarities than differences. The only clear difference between these two enzymes was related to the hydrolysis of LH-RH and substance P. The recombinant rat testes EOP 24.15 hydrolyzed the LH-RH and the substance P which were not substrates for rabbit brain EOPA. The resistance of LH-RH to hydrolysis by the rabbit brain enzyme was recently confirmed by Lew *et al.* (1994) using the enzyme obtained from ovine brain. Another alleged difference between EOPA and EOP 24.15 is related to the hydrolysis of angiotensin II. This difference was not confirmed in the experiments presented here. Similarly to our previous result (Camargo *et al.*, 1979b) and in contrast to the results reported by Chu and Orlowski (1985), angiotensin II was not hydrolyzed by the recombinant testes EOP 24.15. It is possible that the hydrolysis of angiotensin II by the rat brain EOP 24.15 referred by Chu and Orlowski (1985) was due to the presence of co-purified peptidases such as the soluble-angiotensin II binding protein, another EOP 24.15-related protease (Kato *et al.*, 1994) which is able to degrade angiotensin II (Kawabata *et al.*, 1993).

Several oligopeptidases have been identified and characterized which display the same specificity towards enkephalin-containing peptides (for review see Barrett *et al.*, 1995). For example, the neurolysin (EC 3.4.24.16) which is also found in nervous tissue and is 65% homologous to EOP 24.15, displays the same specificity toward opioid peptides as the EOP 24.15 (Dauch *et al.*, 1995). Thus the application of the internally quenched fluorescence opioid derived substrate (QF-ERP$_7$) to measure the EOPA-like activity in crude enzyme tissue preparations is likely to represent the activities of more than a single enzyme. Therefore, the contribution that each particular enzyme makes to the overall cytosolic activity depends on the relative concentration of each enzyme which could vary from tissue to tissue and from one species to another. In conclusion, the lack of correspondence of the EOPA-like activity of cytosol of rat or rabbit brain and the levels of EOP 24.15 mRNA and protein detected in the same subcellular fraction of rat or rabbit brain lead us to suggest that the enzyme responsible for most activity toward QF-ERP$_7$ in the brain of rat or rabbit is the EOPA and not the EOP 24.15.

Except for the differences discussed herein, the rabbit brain EOPA and the rat recombinant EOP 24.15 were strikingly similar both in their physicochemical properties (Carvalho and Camargo, 1981; Orlowski *et al.*, 1983) and in their specificities towards bradykinin, neurotensin and metorphinamide (Table 1). This is also true with respect to the effect of EDTA which originally had been the basis for the distinction between EOPA and EOP 24.15. As shown in Figure 4, the recombinant rat testes EOP 24.15 was not inhibited by an increasing concentration of this chelating compound (up to 20 mM). However, the primary amino acid sequence deduced from the cloned rat testes EOP 24.15 cDNA showed the presence of the metallo-endopeptidase motif HEXXH (Pierotti *et al.*, 1990), which would suggest the presence of one catalytic metal coordinated by one water molecule, by the two histidines and by the glutamic acid (see Vallee and Auld, 1990). The metal analysis of the recombinant EOP 24.15 by atomic absorption spectroscopy indicated the presence of 1.8 mol of Zn / mol of enzyme, thus suggesting that at least one Zn would be structural and coordinated by the cysteines which are not involved in forming disulfide bonds (Barrett *et al.*, 1995). If, in fact, EOP 24.15 is a metallo-protease, it should be included among the metallo-enzymes exhibiting low dissociation of the metal (Stöcker *et al.*, 1988), thus explaining its resistance to chelating compounds. However, a direct evidence for the involvement of Zn in the catalytic mechanism of EOP 24.15 is still necessary, since the presence of a catalytic Zn binding motif can also be found in non-proteolytic enzymes such as the factor XII of blood clotting (Bernardo *et al.*, 1993).

The effect of dithiothreitol demonstrated that both the rabbit brain and the rat testes enzymes are thiol-activated proteases, allowing their distinction from neurolysin, an enzyme of the EOP 24.15 family (Dauch *et al.*, 1995). The enzyme activation may be due to the effect of this reducing agent on a critical cysteine which is likely to be in the vicinity of the active center of EOP 24.15. The presence of a critical cysteine derived from the observation of the rapid irreversible inactivation of EOP 24.15 by YGGFRRC[Npys]R(NH)$_2$, in contrast to the slow enzyme inhibition by Cys-Npys. These results are similar to the effect of the Cys(Npys) dynorphin-related compounds toward EOPA (Gomes *et al.*, 1993), thus suggesting that the reactivity of these compounds depend on the presence of

the Cys(Npys) group within the structure of a peptide substrate. The reactivation of the fully inactivated enzyme by DTT also suggests the involvement of a cysteine in the mechanism of inhibition by YGGFRRC[Npys]R$(NH)_2$. Moreover, in analogy to the brain EOPA (Gomes *et al.*, 1993), the inhibition by YGGFRRC[Npys]R$(NH)_2$ was protected by the presence of a substrate such as bradykinin (data not shown), thus reinforcing the hypothesis that the reactive cysteine may be in the vicinity of the active center of the enzyme or may participate at the catalysis.

We conclude that

(i) the level of the EOP 24.15 does not account for the EOPA-like activity in rat brain and in rabbit brain and testes;

(ii) the predominant enzyme in rat brain which displays the peculiar specificity of the EOPA toward neuropeptides is likely to be the rat version of the EOPA and not rat testes EOP 24.15.

Yet, the number of similarities between these enzymes favor the hypothesis that they are two distinct members of the EOP 24.15 family. This hypothesis is corroborated by the fact that other enzymes, structurally related to the rat testes EOP 24.15, have been described in brain (Papastoitsis *et al.*, 1994; Dauch *et al.*, 1995) and in different peripheral tissues (Serizawa *et al.*, 1995; Kato *et al.*, 1994; Kawabata *et al.*, 1993; Rodd *et al.*, 1995).

Materials and Methods

Materials

Rabbit brains, stored at – 20 °C, were obtained from a local slaughter house. The internally quenched fluorescent enkephalin-related peptide (QF-ERP$_7$) was synthesized by classical solution method (Juliano *et al.*, 1990). The Cys(Npys)-compounds were prepared by conventional solid phase method and were generous gifts of Dr. Rei Matsueda (New Lead Research Laboratories, Sankyo Co., Ltd., Shinagawa-ku, Tokyo,140, Japan). Bradykinin, neurotensin, dynorphin A_{1-17}, metorphinamide, substance P, LH-RH, angiotensin II, BAM-22-P and β-endorphin were from Peninsula Laboratories. Succinyl-bis-bradykinin (suc.bis bradykinin) was a generous gift of Dr. Raymond J. Vavrek (Department of Biochemistry, University of Colorado Medical School, Denver). All peptides were subjected to purification by preparative reverse phase HPLC and the amino acid composition and concentrations were determined by pico Tag amino acid analysis after acid hydrolysis. Nylon membranes and the UV Crosslinker RPN2501 were from Amersham Buchler; Taq DNA polymerase and pre-packed oligo-dT cellulose columns were from GIBCO Laboratories; bovine serum albumin was from Sigma Chemical Co.; reagents for Westernblot and the RiboClone cDNA synthesis system were from Promega Biotech.; pGEX was from Pharmacia-LKB; Centricon 50 and μBondapak column were from Millipore Corp.; Vectastain was from Vector Labs.; nitrocellulose membranes (pore size 0.2 μm) were from Schleicher & Schuell.

Preparation of Cytosol of Rat and Rabbit Brain and Testes and Purification of Rabbit Brain EOPA

Fresh testes and brain obtained from mature rat (Wistar) and rabbit (New Zealand) were homogenized in 1:3 weight: volume of 10 mM Tris-HCl buffer pH 7.5, containing 0.25 M sucrose and centrifuged at 25 000 *g* for 15 min. The supernatant fractions were further centrifuged at 100 000 *g* for 1 h and the resulting supernatant fractions were used to determine the enzyme activity. Endo-oligopeptidase A was purified to apparent homogeneity from the cytosol of 250 g of frozen rabbit brains by the procedure involving ammonium sulphate precipitation, molecular sieving, ion-exchange chromatography, and preparative gel electrophoresis (Carvalho and Camargo, 1981). The enzyme obtained from the pooled fractions of preparative gel electrophoresis was introduced into a protein-Pak DEAE-5PW column which was previously equilibrated with 20 mM Tris-HCl buffer, pH 7.5, containing 0.5 mM DTT (Gomes *et al.*, 1993). A 6700-fold purification was achieved with 2.5% recovery of the homogenate activity. SDS-PAGE of the purified rabbit brain EOPA showed a single band corresponding to M_r of 70 000. The specific activity of the enzyme preparation used in the present study was 2130 mU/mg of protein.

Preparation of Recombinant Rat Testes EOP 24.15 and of the Anti-[Recombinant Rat Testes EOP 24.15] Antibody

The purified recombinant rat testes EOP 24.15 was obtained as described by Glucksman and Roberts (1995). Briefly, bacteria containing the EOP 24.15 cDNA inserted in the plasmid pGEX were grown with antibiotic selection to an absorbance reading of 0.6 at 600 nm, when the expression of the EOP 24.15-fusion gene was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG), proceeding for 3 h. Bacteria were lysed by two cycles of freezing and thawing followed by sonication. After removal of bacterial debris, the supernatant was incubated with gluthadione-sepharose beads (GST), and subsequently with thrombin to cleave at the junction of the two genes. Contaminating thrombin was removed by exhaustive filtration with a Centricon 50. The yield was 2.5 mg of pure protein per liter of culture, as assayed by native and SDS-polyacrylamid gel electrophoresis.

The purified recombinant rat testes EOP 24.15 was subjected to acid hydrolysis and the amino acid composition given by the amino acid analysis (Hendrikson and Meredith, 1984) reproduced the theoretical composition values of the EOP 24.15 (Pierotti *et al.*, 1990). The specific activity of the recombinant rat testes EOP 24.15 used in this work was 2.2 U/mg.

For the preparation of the affinity purified antiserum, the enzyme was mixed with complete Freud's adjuvant and injected intradermally into New Zealand white rabbits. Initial injections of 20 μg recombinant EOP 24.15 were used for immunization followed by three subsequent injections of 10 μg of protein at two week intervals. Antiserum was collected, centrifuged and further affinity-purified. The antiserum was precipitated with 50% ammonium sulfate, resuspended in 10 mM sodium phosphate, pH 8.0, and dialysed against 0.1 M phosphate buffer, pH 7.0 (PBS). The dialysate was adjusted to 10% glycerol and passed through a GST-EP 24.15 column. The column was washed with PBS and subsequently with PBS containing 2 M KCl, and bound antibody was eluted with freshly prepared 5 M NaI containing 1 mM sodium thiosulfate. Affinity-purified anti-recombinant rat testis EOP 24.15 antibody was stored at – 20 °C in 50 μl aliquots.

Northern Blot

A single-step method for RNA isolation using acid guanidinium thiocyanate-phenol-chloroform extraction was employed (Puissant and Houdebine, 1990). Total RNA (10 μg) of whole brain or testes of rat and rabbit were submitted to electrophoresis in denaturing agarose gels (1.7% formaldehyde) and transferred to nylon membranes (Ausubel *et al.*, 1994). The mRNA of rabbit brain

and testes was obtained by purification using pre-packed oligo-dT cellulose columns, and 5 μg were electrophoresed like above. The RNA was fixed on the membrane by UV crosslinking. Membranes were pre-hybridized overnight at 42 °C in 50% formamide, 25 mM K_2PO_4, pH 7.4, 5X Denhardts solution, 50 μg/ml herring sperm DNA, 10% dextran sulfate (Ausubel *et al.*, 1994). Hybridizations with the radiolabeled cDNA of recombinant rat testes EOP 24.15 (Pierotti *et al.*, 1990) and the control β-actin were performed for 16 h at 42 °C, adding the probe to the prehybridization solution (1.5×10^6 cpm/ml). The cDNA fragments were radiolabeled with ^{32}P-dATP using the random primer procedure (Harris *et al.*, 1992). The blots were washed using high stringency conditions: four washes at 65 °C in 2 x SSC, 0.1% SDS for 15 min, and three washes at 65 °C in 0.1 x SSC, 0.1% SDS for 10 min. The blots were exposed to X-ray film for 24 h (total RNA blots), or 72 h (mRNA blots). The band intensities were determined using a densitometer.

The rabbit β-actin gene was obtained by PCR using the following oligonucleotides: 5′ ACAGTCCGCCTAGAAGCACTT 3′ (sense) and 5′ CATCCACGAGACCACCTTCAA 3′ (anti-sense) designed following the nucleotide sequence of rabbit β-actin cloned by Harris *et al.* (1992). Double-stranded cDNA of rabbit renal tissue was used as template in PCR. The sequence of rabbit β-actin obtained (319 bp) was confirmed by double-stranded dideoxy chain termination sequencing using vector derived primers.

SDS/PAGE and Immunoblotting

Samples of 2 milliunits of each of recombinant rat testes EOP 24.15 (0.27 μg), of the cytosols of rat brain (0.6 mg) and testes (0.15 mg), purified rabbit brain EOPA (0.21 μg) and M_r standards were subjected to a SDS-PAGE (Laemmli, 1970), and electrotransferred to a nitrocellulose membrane for 18 h at 180 mA using a Bio-Rad Trans-Blot cell (Towbin *et al.*, 1979). The nitrocellulose membrane was subsequently soaked in blot buffer (150 mM NaCl, 1 mM EDTA, 30 mM Tris-HCl buffer, pH 7.3, 3% BSA, 0.05% Tween 20 and 0.05% NaN_3) for 4 h before incubation for 1 h with the 1:500 dilution of anti-[recombinant rat testes EP 24.15] antiserum. The Western blot was developed using an alkaline phosphatase conjugated anti-mouse IgG diluted 1:7500 in 3% BSA for 30 min and reacted with BCIP/NBT following the procedures described by the manufacturers. Control experiments were conducted using Balb-c mouse pre-immune antiserum.

Enzyme Assays

The fluorogenic substrate QF-ERP_7 was used to determine EOPA and recombinant rat testes EOP 24.15 activities, as previously described (Juliano *et al.*, 1990). One unit of the EOPA-like activity is the amount of enzyme which hydrolyses 1 μmol of QF-ERP_7 in 1 min. All the enzymatic assays were performed in triplicate.

Inhibition of the Recombinant Rat Testes EOP 24.15 Activity by YGGFLRRC(Npys)R(NH)2 and by Cys(Npys)

The irreversible inhibition by Cys(Npys) compounds was determined according to Gomes *et al.* (1993) using 2 mU of EOP 24.15 (approximately 1 μg of protein), pre-incubated at 37 °C, with 10 μM of the inhibitors in 100 μl of 20 mM Tris-HCl buffer, pH 7.5. Aliquots of 10 μl of the enzyme-inhibitor pre-incubation solution were removed at time intervals (1, 5, 30 and 60 min) and diluted in 500 μl of the same buffer containing 16 μM of QF-ERP_7 and assayed fluorimetrically at 37 °C as described (Juliano *et al.*, 1990). Bradykinin protection experiments were performed by addition of YGGFLRRC(Npys)R$(NH)_2$ to 100 μl of the pre-incubation solution containing 2 mU of enzyme and 20 μM of bradykinin. The effect of DTT was performed by addition of 0.5 mM of DTT to the enzyme which had been pre-incubated for 1 min with YGGFLRRC(Npys)R$(NH)_2$. All the assays were performed in triplicate.

Determination of the Hydrolysis Rates of P ptides by HPLC

The rates of hydrolysis of peptides were determined with digestion mixtures containing 30 – 50 μM substrate and 1.5 mU of enzyme in 20 mM Tris-HCl buffer, pH 7.5, containing 0.5 mM DTT. The reactions were carried out at 37 °C and terminated by the addition of 10 μl of 10% (v/v) trifluoroacetic acid. Peptide fragments were separated by reverse phase HPLC using a C18 μBondapak column (4.6 x 250 mm) with a linear gradient of 0 – 35% acetonitrile in 0.1% trifluoroacetic acid in 15 min at a flow rate of 2 ml/min. Absorbance was monitored at 214 nm.

Identification of Cleavage Sites in Peptides

Cleavage sites of peptides were identified by determining the amino acid composition of each product generated from the incubation with enzymes and separated by HPLC. Peptide fragments generated by recombinant rat testes EOP 24.15 or rabbit brain EOPA were separated by reverse phase HPLC using a C_{18} μBondapak column. Peptides were eluted with a linear gradient of 0 – 35% acetonitrile in 0.1% trifluoroacetic acid in 15 min at a flow rate of 2 ml/min. Absorbance was monitored at 214 nm. The isolated peptides were lyophilized, hydrolyzed for 22 h at 110 °C in 6 M HCl, containing 1% phenol, in tubes sealed in vacuum and then subjected to amino acid analysis using a pico Tag station (Hendrikson and Meredith, 1984).

Protein concentration

Protein concentration determinations from cytosols of tissues were performed by the method of Lowry *et al.* (1951). The purified enzymes were lyophilized, hydrolyzed for 22 h at 110 °C in 6 M HCl containing 1% phenol in tubes sealed in vacuum. The protein concentrations were deduced after amino acid analysis using a pico Tag station (Hendrikson and Meredith, 1984).

Acknowledgements

We are grateful to Drs. Marc Glucksman and James Roberts from the Mount Sinai School of Medicine, New York, USA, for providing the recombinant plasmid containing the cDNA clone of endo-oligopeptidase 24.15. We thank Dr. Rei Matsueda, Sankyo Co., Tokyo, Japan, for the Cys(Npys)-derived compounds. We are also grateful to Mauro Sucupira for technical assistance. This work was supported by Fundação de Amparo e Pesquisa do Estado de São Paulo.

References

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994). Preparation and analysis of RNA. In: Current Protocols in Molecular Biology, Suppl. *21*, 4.9.7. (New York: Wiley Interscience).

Barrett, A.J., and Rawlings, N. D. (1992). Oligopeptidases, and the emergence of the prolyl-oligopeptidase family. Biol. Chem. Hoppe-Seyler *373*, 353 – 360.

Barrett, A. J., Brown, M. A., Dando, P. M, Knight, C. G., McKie, N., Rawlings, N. D., and Serizawa, A. (1995). Thimet oligopepti-

dase and oligopeptidase M or neurolysin. Methods in Enzymol. *248*, 529-554.

Bernardo, M.M., Day, E.D., Halvorson, H.R., Olson, S.T., and Shore, J.D. (1993). Surface-independent acceleration of factor XII activation by Zn ion. J. Biol. Chem. *268*, 12477 – 12483.

Camargo, A.C.M., Shapanka, R., and Greene, L.J. (1973). Preparation, assay and partial characterization of a neutral endopeptidase from rabbit brain. Biochemistry *12*, 1838 – 1844.

Camargo, A.C.M., Reis, M.L., and Caldo, H. (1979a). Susceptibility of a peptide derived from bradykinin to hydrolysis by brain endo-oligopeptidases and pancreatic proteinases. J. Biol. Chem. *254*, 5304 – 5307.

Camargo, A.C.M., Martins, A.R., and Greene, L.J. (1979b). Steric constraints make polypeptides resistant to hydrolysis by tissue peptidases. In: Limited Proteolysis in Microorganisms, DHEW Publication n° (NIH)78-1591, G.N. Cohen and H. Holzer, eds. (Washington, D.C.: U.S. Government Printing Office), pp. 45 – 48.

Camargo, A.C.M., Fonseca, M.J.V, Caldo, H., and Carvalho, K.M. (1982). Influence of the carboxyl terminus of Luteinizing Hormone-Releasing Hormone and Bradykinin on hydrolysis by brain endo-oligopeptidases. J. Biol. Chem. *257*, 9265 – 9267.

Camargo, A.C.M., Caldo, H., and Emson, P.C. (1983). Degradation of neurotensin by rabbit brain endo-oligopeptidase A and endo-oligopeptidase B (Proline-endopeptidase). Biochem. Biophys. Res. Commun. *166*, 1151 – 1159.

Camargo, A.C.M., Ribeiro, M.J.V.F., and Schwartz, W. (1985). Conversion and inactivation of opioid peptides by rabbit brain endo-oligopeptidase A. Biochem. Biophys. Res. Commun. *130*, 932 – 938.

Camargo, A.C.M., Oliveira, E.B , Toffoletto, O., Metters, K.M., and Rossier, J. (1987). Brain endo-oligopeptidase A, a putative enkephalin-converting enzyme. J. Neurochem. *48*, 1258 – 1263.

Camargo, A.C.M., Gomes, M.D., Toffoletto, O., Ribeiro, M.J.F., Ferro, E.S., Fernandes, B.L., Suzuki, K. , Sasaki, Y., and Juliano, L. (1994). Structural requirements of bioactive peptides for the interaction with endopeptidase 22.19. Neuropeptides *26*, 281 – 287.

Carvalho, K.M., and Camargo, A.C.M. (1981). Purification of rabbit brain endo-oligopeptidases and preparation of anti-enzyme antibodies. Biochemistry *20*, 7082 – 7088.

Chu, T.G., and Orlowski, M. (1985). Soluble metalloendopeptidase from rat brain: action on enkephalin-containing peptides and other bioactive peptides. Endocrinology *116*, 1418 – 1425.

Collins, J.F., Honda, T., Knobel, S., Bulus, N.M., Conary, J., DuBois, R., and Ghishan, F.K. (1993). Molecular cloning, sequencing, tissue distribution, and functional expression of a Na+/H+ exchanger (NHE-2). Proc. Natl. Acad. Sci. USA *90*, 3938 – 3942.

Dauch, P., Vincent, J.P., and Checler F. (1995). Molecular cloning and expression of rat brain endopeptidase 3.4.24.16. J. Biol. Chem. *270*, 27266 – 27271.

Enzyme Nomenclature, Recommendations (1989). Eur. J. Biochem. *179*, 518.

Enzyme Nomenclature, Recommendations (1992). Nomenclature Committee, International Union of Biochemistry and Molecular Biology. (Orlando: Academic Press).

Gomes, M.D., Juliano, L., Ferro, E.S., Matsueda, R., and Camargo, A.C.M. (1993). Dynorphin-derived peptides reveal the presence of a critical cysteine for the activity of brain endo-oligopeptidase A. Biochem. Biophys. Res. Commun. *197*, 501 – 507.

Glucksman, M. J., and Roberts, J. L. (1995). Strategies for characterizing, cloning, and expressing soluble endopeptidases. In: Methods in Neurosciences – Peptidases and Neuropeptide Processing, I.A. Smith, ed. (New York: Academic Press), pp. 281 – 285.

Hardie, D.G. (1992). Biochemical messengers. In: Hormones, Neurotransmitters and Growth Factors (London: Chapman and Hall), pp. 1 – 43.

Harris, D.E., Warshaw, D.M., and Periasamy, M. (1992). Nucleotide sequences of the rabbit α-smooth-muscle and β-non-muscle actin mRNAs. Gene *112*, 265 – 266.

Healy, D.P., and Orlowski, M. (1992). Immunocytochemical localization of endopeptidase 24.15 in rat brain. Brain Res. *571*, 121 – 128.

Henrikson, R.L., and Meredith, S.C. (1984). Amino acid analysis by reverse-phase high-performance liquid chromatography: precolumn derivatization with phenylisothiocyanate. Anal. Biochem. *436*, 65 – 74.

Juliano, L., Chagas, J.R., Hirata, I.Y., Carmona, E., Sucupira, M., Oliveira, E.S., Oliveira, E.B., and Camargo A.C.M. (1990). A selective assay for endooligopeptidase A based on the clivage of fluorogenic substrate structurally related to enkephalin. Biochem. Biophys. Res. Commun. *173*, 647 – 652.

Kato, A., Sugiura, N., Hagiwara, H., and Hirose, S. (1994). Cloning, amino acid sequence and tissue distribution of porcine thimet oligopeptidase. Eur. J. Biochem. *221*, 159 – 165.

Kawabata, S., Nakagawa, K., Muta, T., Iwanaga, S., and Davie, E.W. (1993). Rabbit liver microsomal endopeptidase with substrate specificity for processing proproteins is structurally related to rat testes metalloendopeptidase 24.15. J. Biol. Chem. *268*, 12498 – 12503.

Laemmli, U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature *227*, 680 – 685.

Lasdun, A., Reznik, S., Molineaux, C.J., and Orlowski, M. (1989). Inhibition of endopeptidase 24.15 slows the in vivo degradation of Luteinizing Hormone-Releasing Hormone. J. Pharmacol. Exp. Therap. *251*, 439 – 447.

Lew, R.A., Tetaz, T.J., Glucksman, M.J. , Roberts, J.L., and Smith, I. (1994). Evidence for a two-step mechanism of gonadotropin-releasing hormone metabolism by prolyl endopeptidase and metalloendopeptidase EC 3.4.14.15 in ovine hypothalamic extracts. J. Biol. Chem. *269*, 12626 – 12632.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. J. Biol. Chem. *193*, 265 – 275.

Nakagawa, T., Okano, H., Furuichi, T., Aruga, J., and Mikoshiba, K. (1991). The subtype of the mouse inositol 1,4,5,-triphosphate receptor are expressed in a tissue-specific and developmentally specific manner. Proc. Natl. Aca. Sci. USA *88*, 6244 – 6248.

Oliveira, E.B., Martins, A.R., and Camargo, A.C.M. (1976). Isolation of brain endopeptidases: Influence of size and sequence of substrates structurally related to bradykinin. Biochemistry *16*, 1967 – 1974.

Oliveira, E.S., Leite, P.E., Spillatini, M.G., Camargo, A.C.M., and Hunt, S.P. (1990). Localization of endo-oligopeptidase (EC 3.4.22.19) in the rat nervous tissue. J. Neurochem. *55*, 1114 – 1121.

Orlowski, M., Michaud, C., and Chu, T.G. (1983). A soluble metalloendopeptidase from rat brain. Eur. J. Biochem. *135*, 81 – 88.

Orlowski, M., Reznik, S., Ayala, J., and Pierott, A.R. (1989). Endopeptidase 24.15 from rat testes. Biochem. J. *261*, 951 – 958.

Orlowski, J., Kandasamy, R., and Shull, G.E. (1992). Molecular cloning of putative members of the Na+/H+ exchanger gene family. J. Biol. Chem. *267*, 9331 – 9339.

Papastoitsis, G., Siman, R., Scott, R., and Abraham, C.R. (1994). Identification of a metalloprotease from Alzheimer's disease

brain able to degrade the β-amyloid precursor protein and generate amyloidogenic fragments. Biochemistry *29*, 10323 – 10329.

Park, J., Freedman, R., Bach, C., Yee, C., Rohrwild, M., Kaminishi, H., Muller Esterl, W., and Jarnagin, K. (1994). Bradykinin-B2 receptors in humans and rats: cDNA structures, gene structure, possible alternative splicing, and homology searching for subtypes. Braz. J. Med. Biol. Res. *27*, 1707 – 1724.

Piechaczyk, M., Blanchard, J.M., Marty, L., Dani, C., Panabieres, F., El Sabouty, S., Fort, P., and Jeanteur, P. (1984). Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. Nucl. Ac. Res. *12*, 6951 – 6963.

Pierotti, A., Dong, K.W., Glucksman, M.J., Orlowski, M., and Roberts, J.L. (1990). Molecular cloning and primary structure of rat testes metalloendopeptidase EC 3.4.24.15. Biochemistry *29*, 10323 – 10329.

Puissant, C., and Houdebine, L. (1990). An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. BioTechniques *8*, 148 – 149.

Rodd, D., and Hersh, L.B. (1995). Endopeptidase 24.16 B. J. Biol. Chem. *270*, 10056 – 10061.

Serizawa, A., Dando, P.M., and Barrett, A.J. (1995). Characterization of a mitochondrial metallopeptidase reveals neurolysin as a homologue of thimet oligopeptidase. J. Biol. Chem. *270*, 2092 – 2098.

Shaw E. (1990). Cysteinyl proteinases and their selective inactivation. Advances in Enzymology *63*, 271 – 347.

Shields, A., and Wickramasinghe, S.N. (1995). Expression of an erythropoietin-like gene in the trout. Br. J. Haematol. *90*, 219 – 221.

Stöcker W., Wolz, R.L., and Zwinlling, R. (1988). Astacus protease, a zinc metalloenzyme. Biochemistry *27*, 5026 – 5032.

Toffoletto, O., Metters, K.M., Oliveira, E.B., Rossier J., and Camargo, A.C.M. (1988). Enkephalin is liberated from [Met^5] enkephalin-R-R-V-NH2 and dynorphin A_{1-8} by endo-oligopeptidase A but not by Metalloendopetidase EC.3.4.24.15. Biochem. J. *252*, 35 – 38, 1988.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA *76*, 4350 – 4354.

Vallee, B.L., and Auld, D.S. (1990). Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry *29*, 5647 – 5659.

Received January 3, 1996; accepted May 1, 1996

Biochemical and Biophysical Research Communications 269, 7–13 (2000)
doi:10.1006/bbrc.2000.2243, available online at http://www.idealibrary.com on IDEAL®

Molecular and Immunochemical Evidences Demonstrate That Endooligopeptidase A Is the Predominant Cytosolic Oligopeptidase of Rabbit Brain

Mirian A. F. Hayashi,* Fernanda C. V. Portaro,* Denise V. Tambourgi,† Mauro Sucupira,‡ Tetsuo Yamane,* Beatriz L. Fernandes,§ Emer S. Ferro,¶ Nancy A. Rebouças,‖ and Antonio C. M. de Camargo*

*Department of Biophysics and Biochemistry and †Department of Immunology, Butantan Institute, São Paulo, Brazil; ‡Department of Pharmacology, §Department of Microbiology, ¶Department of Histology and Embryology, and ‖Department of Biophysics and Physiology, Biomedical Sciences Institute, University of São Paulo, São Paulo, Brazil

Received December 29, 1999

Oligopeptidases are tissue endopeptidases that do not attack proteins and are likely to be involved in the maturation and degradation of peptide hormones and neuropeptides. The rabbit brain endooligopeptidase A and the rat testes soluble metallopeptidase (EC 3.4.24.15) are thiol-activated oligopeptidases which are able to generate enkephalin from a number of opiod peptides and to inactivate bradykinin and neurotensin by hydrolyzing the same peptide bonds. A monospecific antibody raised against the purified rabbit brain endooligopeptidase A allowed the identification of a 2.3 kb cDNA coding for a truncated enzyme of 512 amino acids, displaying the same enzymatic features as endooligopeptidase A. In spite of all efforts, employing several strategies, the full-length cDNA ould not be cloned until now. The analysis of the deduced amino acid sequence showed no similarity to the rat testes metalloendopeptidase sequence, except for the presence of the typical metalloprotease consensus sequence [HEXXH]. The antibody raised against recombinant endooligopeptidase A specifically inhibited its own activity and reduced the thiol-activated oligopeptidase activity of rabbit brain cytosol to less than 30%. Analysis of the endooligopeptidase A tissue distribution indicated that this enzyme is mainly expressed in the CNS, whereas the soluble metallo EC 3.4.24.15 is mainly expressed in peripheral tissues.

Neuropeptides, displaying a multiplicity of neurochemical functions, have been found throughout the central nervous system (CNS). Modulation of their specific actions depends upon the activity of proteolytic enzymes present in the nervous tissues. These enzymes may cause the complete inactivation and/or conversion of neuropeptides into smaller, active homologues. This is the case for the thiol-activated endooligopeptidase A (EOPA) and the rat testes soluble metalloendopeptidase (EC 3.4.24.15 or EP 24.15), that effectively degrade bradykinin and neurotensin (1–3). Both enzymes convert enkephalin-containing peptides into enkephalins (4, 3) but only EP 24.15 generates a pentapeptide from the N-terminus of GnRH (5, 3), which is a potent inhibitor of the GnRH secretion (6). Based exclusively on the striking similarity between enzymatic properties of rabbit brain EOPA and rat testes EP 24.15, it was suggested that they should be the same enzyme (7). However, a few differences in their biochemical and immunochemical properties (8, 9), and in their subcellular localization in the CNS (10, 11), indicated that EOPA and EP 24.15 could not be the same enzyme.

Here, we describe a cDNA isolated after immunoscreening of a rabbit brain cDNA library, which allowed the expression of a truncated protein displaying the same enzymatic features as the purified rabbit brain EOPA. Analysis of the deduced amino acid sequence showed no similarity to the rat testes EP 24.15 sequence, except for the presence of the typical metallo protease consensus sequence HEXXH (12). The antibody raised against this recombinant enzyme (*r*EOPA) specifically inhibited its own activity, and also reduced the EOPA activity of rabbit brain crude cytosol to less than 30%. Analysis of the EOPA tissue distribution

The DNA sequence presented in this work received the Accession No. AF15037 in GenBank.

Abbreviations used: EOPA, endooligopeptidaseA; *r*EOPA, recombinant EOPA-GST-fusion; *r*EP 24.15, recombinant soluble metalloendopeptidase; QF-ERP$_7$, quenched fluorescence enkephalin-related peptide; pER-12 + MA-16, plasmid containing total cloned cDNA.

0006-291X/00 $35.00

indicated that this enzyme is mainly expressed in the CNS.

MATERIALS AND METHODS

Materials. The purified rabbit brain EOPA used in the present study was the same as described in Hayashi *et al.* (9). New Zealand rabbits were obtained from a local slaughterhouse and the male Balb/c mice were obtained from the local animal house. The peptide bradykinin [BK] and neurotensin [NT] were purchased from Peninsula Laboratories. The quenched fluorescent peptide substrate QF-ERP$_7$ [(orto-aminobenzoyl)-Gly^1-Gly^2-Phe^3-Leu^4-Arg^5-Arg^6-Val^7-(N-(2,4,dinitrophenyl-ethylenediamine)] was synthesized by classical solution method (13). The Cys(Npys) dynorphin-derived peptide [Tyr(D)Ala-Gly-Phe-(D)Leu-Cys(Npys)-NH_2] and the CPP-AAY-pAB [N-[1(R,S)-carboxy-3-phenylpropyl]-Ala-Ala-Tyr-(*p*-aminobenzoate)] were generous gifts of Dr. Rei Matsueda (New Lead Research Laboratories, Sankyo Co., Shinagawa-ku, Tokyo, Japan) and Dr. I. A. Smith (Baker Medical Institute, Melbourne, Australia), respectively.

Cloning and sequencing of the cDNA coding for the EOPA. A rabbit brain cDNA library constructed in λgt11 vector (Clontech) was screened after induction with IPTG using the anti-EOPA polyclonal antibody. Positive clones were identified using the ProtoBlot immunoscreening system (Promega) and were analyzed by sequencing the 5′ ends of the inserts. The longest cDNA insert (2 kb) was amplified by PCR using commercially available oligonucleotide primers for λgt11 (Gibco-BRL), and was subcloned into the vector pBluescript SK+ (Stratagene). This clone was named pER-12. Both strands of the cDNA insert were sequenced by the chain terminator dideoxy method (14), using Sequenase kit (USB), and by primer walking sequencing. The "fmol kit" (Promega) and the boiling method of DeShazer *et al.* (15) were employed for sequencing especially difficult regions. Amplification of sequences upstream of the cloned cDNA was performed using the "Marathon cDNA Amplification Kit" (Clontech) on mRNA of rabbit brain, following the manufacturer's instructions. A fragment overlapping with pER-12 (MA-16) was obtained and fused in frame to the 5′ end of the cDNA insert of pER-12, yielding the sequence coding for the truncated EOPA (see below).

Northern blot analysis. 10 μg of total RNA from each tissue of a male New Zealand rabbit were subjected to electrophoresis in denaturing agarose gel (1.7% formaldehyde), transferred to nylon membranes and hybridized as previously described (9). ^{32}P-labeled probes were prepared by random priming (Gibco BRL) using the complete sequence of the pER-12 cDNA insert.

Preparation of crude rabbit brain cytosol, rEP 24.15 and rEOPA. Rabbit brains were homogenized in 1:3 (w/v) of 0.32 M of sucrose solution and the cytosol was obtained by centrifugation at 4°C for 1 h at 100,000 × g (9). Recombinant EP 24.15 (*r*EP 24.15) was obtained by the method described by Glucksman and Roberts (16). The recombinant EOPA (*r*EOPA) was expressed in *Escherichia coli* as a gluthatione-S-transferase (GST) fusion-protein using the pGEX 4T-1 expression vector from Pharmacia (17). After linking the cDNA insert from pER-12 to its 5′ upstream fragment, MA-16, in frame (Fig. 2A), this sequence was inserted into the expression vector pGEX 4T-1 (Pharmacia), downstream of the sequence coding for GST, such that the transcription initiation site of the vector could be used. The expression of the fusion protein was conducted as previously described for the rat testis *r*EP 24.15 (16). The fusion protein was purified after incubation with the glutathione 4B–Sepharose resin (Pharmacia). The homogeneity of the fusion protein was verified by SDS-PAGE (18) and the protein concentration was determined by the method of Bradford (19).

Preparation of antisera. Male Balb/c mice, 7–8 weeks old and weighing 18–22 g, were immunized either with purified EOPA or with *r*EOPA. For each of four immunizations, 2 μg of purified EOPA or 3 μg of the fusion protein absorbed on $Al(OH)_3$ were injected intradermally at weekly intervals. Blood samples were collected one week after the last immunization and the sera were stored at −20°C. The animals were strictly maintained and manipulated under ethical conditions according to the International Animal Welfare Recommendations.

Immunoreaction assays. The anticatalytic activities of the antisera were monitored by HPLC using bradykinin as substrate. Antisera were pretreated at 55°C for 5 min to eliminate any endogenous peptidase activity. In the presence or absence of antisera, 20 μM bradykinin were incubated with 1.5 mU of enzyme in 50 mM Tris–HCl, pH 7.5, 20 mM NaCl, containing 0.5 mM DTT. The reactions were stopped by the addition of 10 μl of 10% TFA (v/v) and the formation of bradykinin products was determined by HPLC, as described below. The reactivity of the anti-EOPA and anti-*r*EOPA antibodies against the rabbit brain cytosol, purified EOPA, *r*EOPA and *r*EP 24.15 were evaluated by ELISA as previously described (20).

Enzyme assays. The fluorogenic substrate QF-ERP$_7$ was used to determine the thiol-activated oligopeptidase activity, as previously described (21, 9). One unit of the enzyme activity is the amount of enzyme which hydrolyses 1 μmol of QF-ERP$_7$ in 1 min. All the enzymatic assays were performed in triplicate.

Determination of the peptide cleavage sites. The sites of cleavage in bradykinin [BK], neurotensin [NT] and in the quenched fluorescent peptide substrate [QF-ERP$_7$] were determined by reverse-phase HPLC (22, 9). The peptides (20 μM) were incubated with 1.5 mU of the *r*EOPA in 50 mM Tris–HCl buffer, pH 7.5, containing 20 mM NaCl and 0.5 mM DTT. The reactions were carried out at 37°C and stopped by the addition of 10 μl of 10% (v/v) TFA. All enzymatic assays were performed in triplicate.

Effect of DTT and active-site directed inhibitors on rEOPA activity. The DTT activation was performed by incubating 1.5 mU of *r*EOPA in the presence [for the assays with CPP-AAY-pAB, 1 μM] (23) or absence of 0.5 mM DTT [for assays with Y(D)AGF(D)LC(Npys)-NH_2, 10 μM] (21) in 50 mM Tris–HCl, pH 7.5, 20 mM NaCl. After 10 min pre-incubation at 37°C with either compound, the inhibition of *r*EOPA was determined by the fluorimetric assay using the fluorogenic peptide QF-ERP$_7$ as substrate, as described above. Aliquots of 50 μl of the enzyme-inhibitor preincubation solutions were diluted in 500 μl of the same buffer containing 2 K_m of QF-ERP$_7$ [K_m = 5.4 μM].

RESULTS

Characterization of the antibody raised against rabbit brain EOPA. The polyclonal antiserum raised against the purified EOPA was able to inhibit more than 70% of the thiol-activated oligopeptidase activity of rabbit brain cytosol, but not of the *r*EP 24.15 (Fig. 1A). In addition it did not show cross-reactivity with *r*EP 24.15 (Fig. 1B). The ability of this antiserum to detect the antigen in the crude cytosol of rabbit brain was evaluated by ELISA, and showed strong immunoreactivity (up to the dilution of 1:16000) (Fig. 1B).

Cloning and sequence analysis of the isolated cDNA. The immunological screening of approximately 10^7 plaques of a rabbit brain cDNA library identified three positive clones, which were shown to correspond to the same cDNA by sequencing (data not shown). Complete sequencing of the longest cDNA insert (2 kb) showed that it was incomplete. The use of RT-PCR, 5′-RACE, cloning of 5 prime extension products, or either the screening of other rabbit brain cDNA libraries by hy-

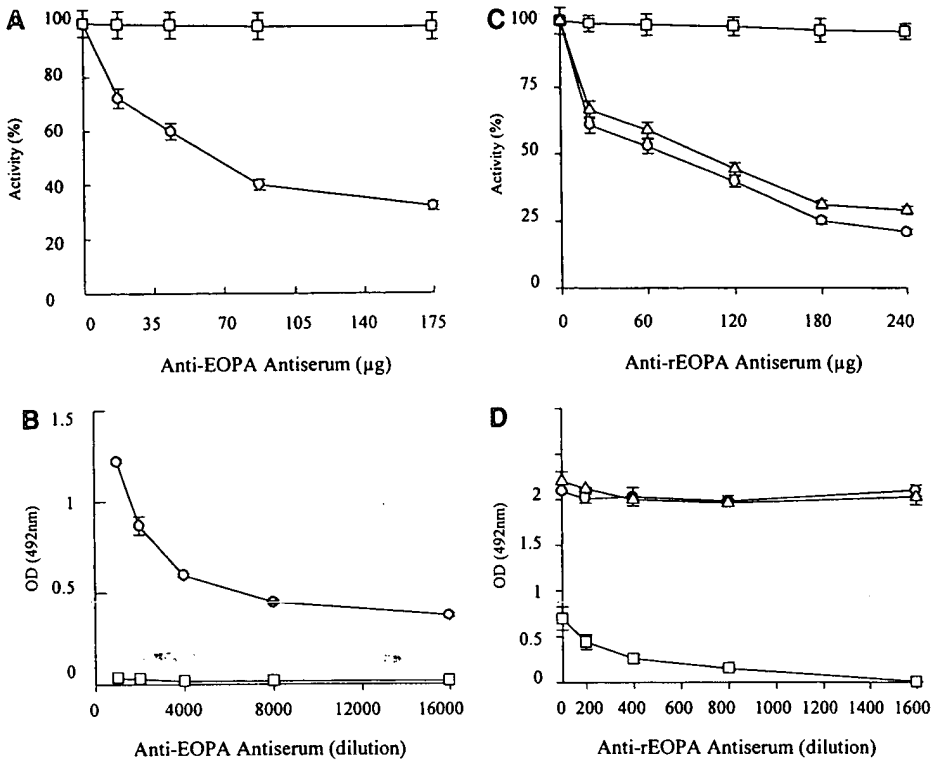


FIG. 1. Immunoreactivity of the anti-EOPA and the anti-*r*EOPA antisera. (A) The anticatalytic activity of anti-EOPA antiserum toward *r*EP 24.15 (□) and the thiol-activated oligopeptidase activity of rabbit brain cytosol (○). (B) The immunoreactivity detected by ELISA using the crude rabbit brain cytosol (○) and *r*EP 24.15 (□) as antigen sources. (C) The anticatalytic activity of the anti-*r*EOPA antiserum toward *r*EP 24.15 (□), *r*EOPA (△) and the thiol-activated oligopeptidase activity of rabbit brain cytosol (○). (D) The immunoreactivity detected by ELISA using *r*EOPA (△), the crude rabbit brain cytosol (○), and *r*EP 24.15 (□) as antigen sources. In A and C, data are presented as percent of enzymatic activity versus antibody concentration in incubation media.

bridization did not succeed in the cloning of the full-length cDNA. A fragment overlapping with the pER-12 cDNA insert (clone MA-16) was obtained only by using the "Marathon" system.

The total cloned cDNA sequence (pER-12 + MA-16) predicted an open reading frame of 1537 nucleotides which could code for a protein of 512 amino acids (Fig. 2B). The 3′ untranslated region was 755 bp long and presented a single AATAAA polyadenylation signal (24), 6 nucleotides upstream of the last residue. No poly(A)-tail was found. Neither the 5′ untranslated region nor the initial methionine (25) could be identified. Further several attempts at cloning the 5′ end with the "Marathon" system did not generate sequences which reached further upstream from clone MA-16. A human brain cDNA library was screened using the 2.3 kb cDNA (pER-12 + MA-16) as a probe, and the longest cDNA insert isolated (about 1.7 kb) was about 65% similar to the rabbit cDNA sequence but also lacked the 5′ end (data not shown).

Analysis of the deduced amino acid sequence of the rabbit brain total cloned cDNA sequence (pER-12 + MA-16) showed the presence of the typical metalloprotease consensus sequence [HEXXH] at the C-terminus, and a high content of cysteines (18 residues, 3.5%) and prolines (35 residues, 6.8%). Three putative phosphorylation sites for MAP kinase were also identified (Thr_{274}, Thr_{300} and Thr_{416}).

The hydropathic profile (26) showed no hydrophobic segment, typical of transmembrane domains or a putative signal peptide (data not shown).

Searches for nucleotide sequence homology identified a single sequence registered as *Homo sapiens* clone 23596, showing about 83% similarity for a region of about 1430 nucleotides. The function of this putative human protein is unknown. No significant homology with any known protein sequence was found for the deduced amino acid sequence, except for a segment of about 290 amino acid residues (positions 58–332 of the deduced amino acid sequence), which showed 70% similarity to a mitotic phosphoprotein identified in *Xenopus laevis* (27), whose function has still to be established. However, at the level of nucleotide sequences only very little similarity was observed between these two genes.

Northern blot analysis of rabbit tissues. The presence of mRNA corresponding to the pER-12 cDNA insert in different rabbit tissues was analyzed by Northern blot. A single mRNA transcript of approximately

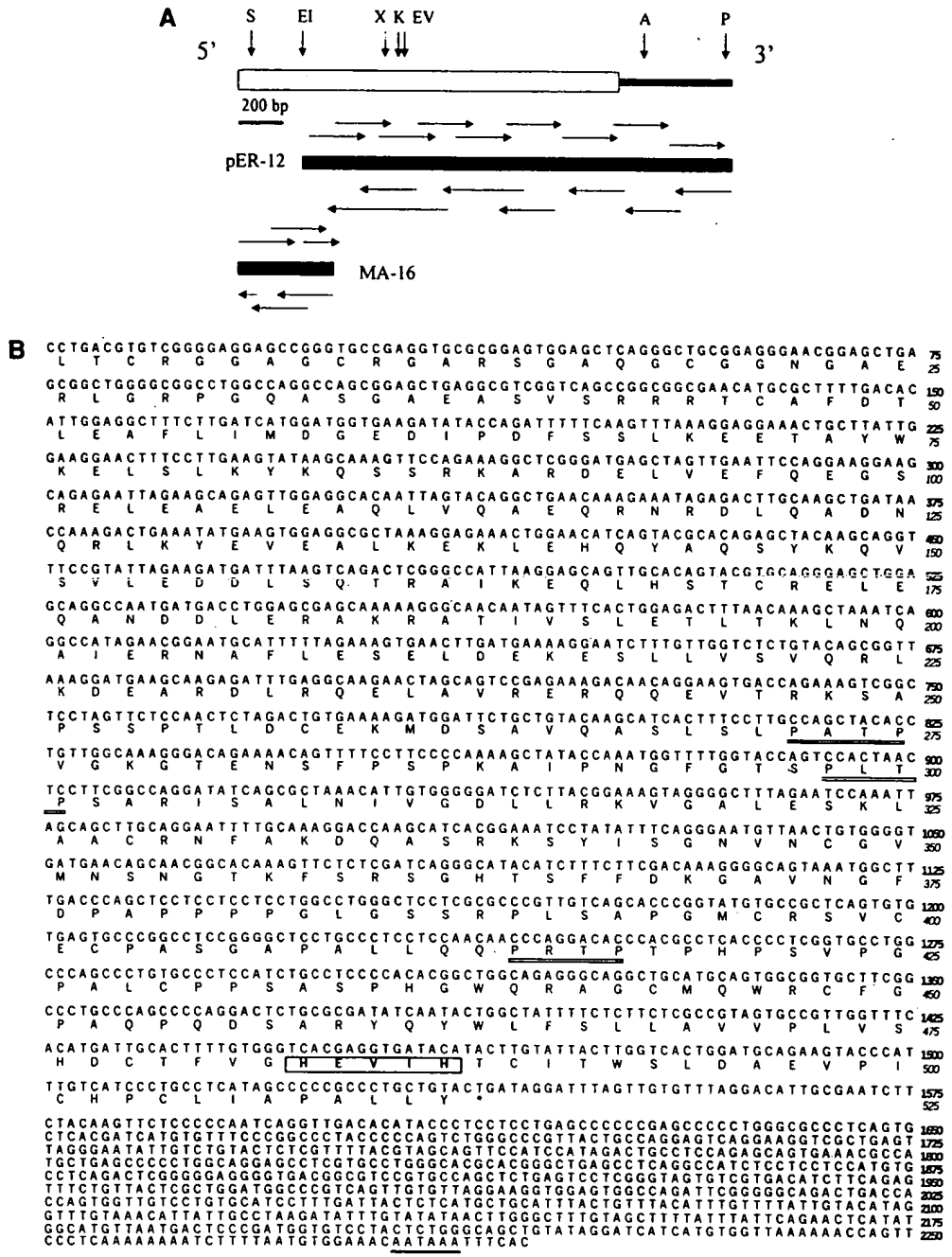


FIG. 2. Schematic representation of the isolated cDNA clones, their sequencing strategy, and the nucleotide and deduced amino acid sequences. (A) A solid line represents the 3′-untranslated region and an open bar indicates the open reading frame. The positions of some restriction enzymes are indicated. The entire sequenced cDNA was reconstituted from two independent overlapping clones, pER-12 and MA-16, as described in the text. Horizontal arrows indicate the direction and the extent of the sequences determined using internal oligonucleotides [EV, *Eco*RV; S, *Sac*I; EI, *Eco*RI; X, *Xba*I; A, *Apa*I; and P, *Pvu*II]. (B) Nucleotides and amino acid residues are numbered in the right column. Amino acids are numbered beginning at the first deduced residue and identified by the single letter code. Three putative MAP kinase phosphorylation sites are doubly underlined and the consensus sequence of metallopeptidases is boxed. An asterisk indicates the stop codon (TGA) of the open reading frame, and the polyadenylation site (AATAAA) is underlined with a single line.

2.8 kb was detected in the total RNA extracted from various rabbit tissues, such as testes, spleen, lung, heart, liver, stomach, hypothalamus, cortex, cerebellum, cerebral stem and striatum (Figs. 3A and 3B). Quantification of the signal intensities showed a clear predominance of transcripts in most of the brain regions, showing a weaker signal in the hypothalamus. In peripheral tissues, the highest abundance was observed in the testes, presenting much lower expression in other tissues (Fig. 3C).

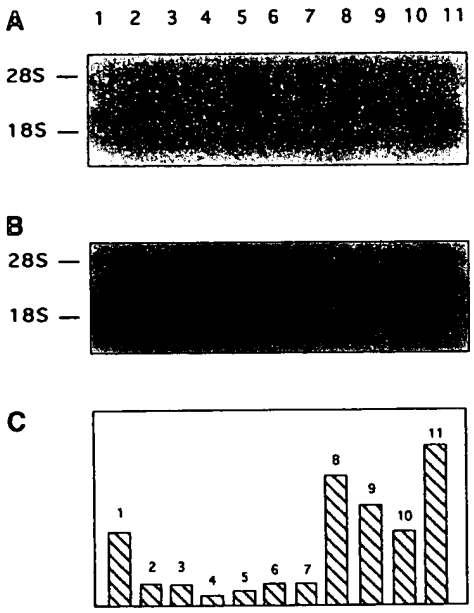


FIG. 3. Tissue distribution of the EOPA mRNA determined by Northern blot analysis. Total RNA (10 μg) from various rabbit tissues were hybridized with the radioactive pER-12 insert, as described under Materials and Methods. The analyzed tissues were: (1) testis, (2) spleen, (3) lung, (4) heart, (5) liver, (6) stomach, (7) hypothalamus, (8) cortex, (9) cerebellum, (10) cerebral stem and (11) striatum. (A) X-ray film after 24 h exposure, (B) ribosomal RNAs from each tissue in agarose gel stained with ethidium bromide and (C) relative intensity of each band quantified by densitometry of the X-ray film. The positions of ribosomal RNAs (28S and 18S) are indicated.

In a different set of analyses, hybridizations performed with total RNA from whole brain and testes of rabbit and rat confirmed the identification of a single band of about 2.8 kb (Fig. 4), as described above. In the testes of both animals, a new band of 1.5 kb was observed besides the 2.8 kb mRNA, suggesting the expression of a homologous mRNA in the testes of both animals. On the other hand, only a single band could be observed when the same membranes were hybridized with EP 24.15 cDNA probes (9). The intensity of the 2.8 kb bands, determined by densitometry, was 2.6 and 1.5 times stronger for the brain as compared to the testes for rabbit and rat, respectively.

Expression of the recombinant protein and enzyme characterization. Expression of the rabbit brain cDNA (pER-12 + MA-16) in *E. coli* generated a GST-fusion protein of about 81 kDa (*r*EOPA), from which the recombinant protein (55 kDa) could be released by digestion with thrombin. Both, the GST-fusion protein and the recombinant protein showed a typical pattern of EOPA/EP 24.15 specificity toward the fluorogenic substrate (QF-ERP$_7$), neurotensin and bradykinin. In fact, a single peptide bond, the Leu^4-Arg^5 of the QF-ERP$_7$, the Arg^8-Arg^9 of neurotensin and the Phe^5-Ser^6 of bradykinin, was hydrolyzed by the *r*EOPA showing a stoichiometric cleavage of these substrates (data not shown). The activity of the *r*EOPA was very unstable, but it could be partially stabilized by the addition of 5% glycerol to the buffer just after its elution from the gluthatione-4B-Sepharose resin.

The *r*EOPA was activated (95.5%) in the presence of low concentrations of DTT (0.5 mM) and was inhibited by either the Cys(Npys) dynorphin-related peptide (about 90% of inhibition with 10 μM) or the CPP-AAY-pAB inhibitor (about 85% of inhibition with 1 μM), similarly as previously observed for purified rabbit brain EOPA (9, 28).

Characterization of immune sera raised against the rEOPA. The polyclonal mouse anti-*r*EOPA antiserum was effective in inhibiting more than 70% of the thiol-activated oligopeptidase activity of both the rabbit brain cytosol and of the *r*EOPA (Fig. 1C). This antiserum did not inhibit the *r*EP 24.15 activity. The ELISA assays showed strong immunoreactivity with *r*EOPA as well as with the rabbit brain cytosol, but no cross-reactivity was detected with the *r*EP 24.15 (Fig. 1D).

DISCUSSION

The specific immunoreactivity of the antibody used to screen the rabbit brain cDNA library was the essential property to perform the isolation of the cDNA coding for the thiol-activated metallo-peptidase EOPA. This was assured by the lack of the antibody's cross-reactivity with EP 24.15. Surprisingly, the amino acid sequence deduced from the isolated cDNA showed that EOPA and EP 24.15 are two distinct metallopeptidases, sharing no sequence similarity. No similarity was found with any other known peptidase. This seems to be another typical example of enzymes displaying similar activities, although lacking homologies in their primary structures, as already described for UCH-L3 (ubiquitin C-terminal hydrolases-isozyme L3) and cathepsin B (29). This observation can be ex-

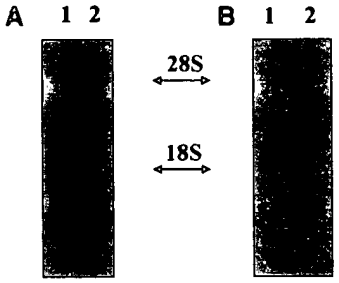


FIG. 4. Comparative Northern blot analysis of rabbit and rat tissues. Total RNA (10 μg) from brain (lane 1) and testes (lane 2) of rabbit (A) and rat (B) were hybridized with the radioactive pER-12 insert, as described under Materials and Methods. The bands observed in the X-ray films are present in (A) and (B). The positions of ribosomal RNAs (28S and 18S) are indicated.

tended to the metalloproteases thermolysin, matrixins, astacins and serralysins (30). Common features in the three dimensional structures of EOPA and EP 24.15 may explain why these two oligopeptidases share similar specificities toward the same substrates and inhibitors (9, 22, 31, 32).

Implying that a single enzyme was responsible for the whole thiol-activated oligopeptidase activity of the CNS, Healy and Orlowski (11) faced difficulties to explain why the EP 24.15 immunoreactivity was restricted to the nucleus of the nervous cells, while most of the thiol-activated oligopeptidase activity had been found in the cytosol (10). Using the anti-rEOPA antibody we clearly demonstrated that EOPA is responsible for over 70% of the thiol-activated neuropeptide-degrading activity of the rabbit brain cytosol. Another important difference between EOPA and EP 24.15 is concerned with the tissue distribution of both oligopeptidases. It was shown here that the EOPA specific mRNA was predominantly found in the CNS, while EP 24.15 mRNA prevailed in peripheral tissues (9, 12). Moreover, recently, Pierotti has shown that EP 24.15 gene is regulated differently in somatic versus germ cells and that the promoter contains a strong negative acting element important for suppressing transcription in non-germ cells, which determine a specific activity 3 to 5 fold higher in testes than in brain (unpublished data).

In spite of the use of several techniques and the screening of different libraries, it was not possible to clone de full-length cDNA coding for endooligopeptidase A. We concluded that the mRNA coding for this enzyme, expressed in both human and rabbit tissues, might present a strong secondary structure imposing the difficulties faced for the cloning of the respective 5′ ends. However, the main purpose of this work was to present structural data to confirm previous suggestions indicating that the oligopeptidase of the cytosol of nervous tissue is distinct from the cytosolic EP 24.15 of peripheral tissues. In fact, the expression of the truncated rabbit brain enzyme (pER-12 + MA-16) allowed us to assure that the enzymatic activity generated was the same as the previously described for the rabbit brain endooligopeptidase A (1–4). Furthermore, the antibody raised against this recombinant enzyme drastically reduced the EOPA activity of rabbit brain crude cytosol. Taken together, these results strongly suggest the existence, in rabbit tissues, of another oligopeptidase beside the soluble metallo-endopeptidase (EP 24.15) sharing similar enzymatic properties. Futhermore, it was also shown that the EOPA-specific mRNA was predominantly found in the CNS, while EP 24.15 mRNA prevailed in peripheral tissues.

The truncated recombinant enzyme (pER-12 + MA-16), displaying the same specificity of the rabbit brain EOPA (4), indicated that the full-length protein was not required to express the active enzyme. This was not surprising since truncated enzymes containing the catalytic domain were found to be able to generate active recombinant proteins (33, 34). Similarly to the natural EOPA and *r*EP 24.15, the fusion protein was activated by the thiol compound (DTT) and inhibited by the metallo- and thiol oligopeptidase inhibitors CPP-AAY-pAB and Y(D)AGF(D)LC(Npys)-NH_2, respectively. Taken together the results presented here demonstrate that not only the (pER-12 + MA-16) cDNA codes for EOPA but also that EOPA is the predominant thiol-activated enzyme in the rabbit brain cytosol able to perform the biotransformation and/or degradation of neuropeptides.

Analyses of the sequence obtained from the clone (pER-12 + MA-16), and its deduced amino acid sequence revealed the presence of a typical metallo-protease consensus motif [HEXXH] at the C-terminus (His_{483}, Glu_{484} and His_{487}). Besides the two histidine ligands of the metallo-protease consensus motif, Vallee and Auld (35) postulated the existence of a third ligand coordinating the catalytic zinc. This ligand is usually a glutamic acid frequently located 13 to 160 residues removed from the consensus sequence. In the case of the EOPA, the putative third ligand could be the glutamic acid residue located 10 residues removed from the second histidine residue of the consensus sequence (Glu_{497}).

The search for similarity in nucleic acid and protein data bases indicated that the (pER-12 + MA-16) cDNA was approximately 83% similar to a region of 1430 nucleotides of a clone from *Homo sapiens* brain coding for a protein with unknown function (36, 37). The deduced amino acid sequence of (pER-12 + MA-16) cDNA showed 70% of similarity for a segment of about 250 amino acid residue of a mitotic phosphoprotein 43 from *Xenopus laevis* (27).

Recently, we showed evidences for the involvement of EP 24.15 in the MHC class I antigen presentation in macrophages (38, 39), a role which might be extended to other peripheral tissues. In CNS, however, the EOPA might have its main physiological role related to the modulation of the action of neuropeptides although no direct evidences have yet been presented. The non-uniform brain distribution of EOPA, which is also a peculiar feature of neuropeptide localization in CNS, certainly favor this hypothesis. The cytosolic localization of EOPA may not limit its involvement in the modulation of neuropeptide metabolism, since it has been demonstrated that EOPA could be secreted by nervous cell (28). Among other possibilities, the enkephalin converting activity of EOPA is one of the most promising hypothesis. Accordingly, the level of EOPA mRNA evaluated by Northern blot analysis was higher in striatum as compared to the brain stem and hypothalamus. Coincidentally, an immunocytochemical survey of EOPA detected a strong immunoreactive "striasome like" structure in the striatum, which cor-

related with enkephalin-rich patches in adjacent sections (10).

ACKNOWLEDGMENTS

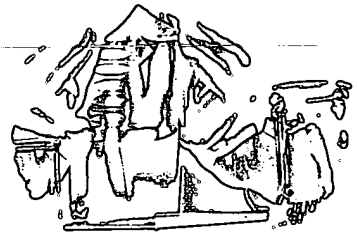
We are grateful to Drs. M. Glucksman and J. Roberts (Mount Sinai School of Medicine, New York) for the EP 24.15 cDNA clone, and we thank Valdeli Braga and Neusa Lima for technical assistance. This work was supported by FAPESP, CNPq, and FINEP.

REFERENCES

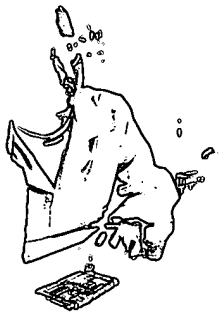
1. Carvalho, K. M., and Camargo A. C. M. (1981) *Biochemistry* **20,** 7082–7088.
2. Camargo, A. C. M., Caldo, H., and Emson, P. C. (1983) *Biochem. Biophys. Res. Commun.* **166,** 1151–1159.
3. Orlowski, M., Reznik, S., Ayala, J., and Pierotti, A. (1989) *Biochem. J.* **261,** 951–958.
4. Camargo, A. C. M., Oliveira, E. B., Toffoletto, O., Metters, K. M., and Rossier, J. (1987) *J. Neurochem.* **48,** 1258–1263.
5. Camargo, A. C. M., Fonseca, M. J. V., Caldo, H., and Carvalho, K. M. (1982) *J. Biol. Chem.* **257,** 9265–9267.
6. Bourguignon, J. P., Alvarez Gonzalez, M. L., Gerard, A., and Franchimont, P. (1994) *Endocrinology* **134,** 1589–1592.
7. Barrett, A. J. (1991) *Biochem. J.* **277,** 295–296.
8. Camargo, A. C. M. (1991) *Biochem. J.* **277,** 294–295.
9. Hayashi, M. A. F., Gomes, M. D., Rebouças, N., Fernandes, B. L., Ferro, E. S., and Camargo, A. C. M. (1996) *Biol. Chem. Hoppe-Seyler* **377,** 283–291.
10. Oliveira, E. S., Leite, P. E., Spillatini, M. G., Camargo, A. C. M., and Hunt, S. P. (1990) *J. Neurochem.* **55,** 1114–1121.
11. Healy, D. P., and Orlowski, M. (1992) *Brain Res.* **571,** 121–128.
12. Pierotti, A., Dong, K. W., Glucksman, M. J., Orlowski, M., and Roberts, J. L. (1990) *Biochemistry* **29,** 10323–10329.
13. Juliano, L., Chagas, J. R., Hirata, I. Y., Carmona, E., Sucupira, M., Oliveira, E. S., Oliveira, E. B., and Camargo, A. C. M. (1990) *Biochem. Biophys. Res. Commun.* **173,** 647–652.
14. Sanger, F., Nicklen, S., and Coulson, A. T. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. DeShazer, D., Wood, G. E., and Friedman, R. L. (1994) *BioTechniques* **17,** 288–289.
16. Glucksman, M. J., and Roberts, J. L. (1995) Methods in Neurosciences—Peptidases and Neuropeptide Processing (I. A. Smith, Ed.), pp. 281–285. Academic Press, New York.
17. Smith, D. B., and Johnson, K. S. (1988) *Gene* **67,** 31–40.
18. Laemmli, U. K. (1970) *Nature* **227,** 680–685.
19. Bradford, M. (1976) *Anal. Biochem.* **72,** 248–254.
20. Tambourgi, D. V., Magnoli, F. C., van den Berg, C. W., Morgan, B. P., Araujo, P. S., Alves, E. W., and Dias da Silva, W. (1998) *Biochem. Biophys. Res. Commun.* **251,** 366–373.
21. Gomes, M. D., Juliano, L., Ferro, E. S., Matsueda, R., and Camargo, A. C. M. (1993) *Biochem. Biophys. Res. Commun.* **197,** 501–507.
22. Camargo, A. C. M., Gomes, M. D., Reichl, A. P., Ferro, E. S., Jacchieri, S., and Juliano, L. (1997) *Biochem. J.* **324,** 517–522.
23. Cardozo, C., and Orlowski, M. (1993) *Peptides* **14,** 1259–1262.
24. Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263,** 211–214.
25. Kozak, M. (1986) *Cell* **44,** 283–292.
26. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
27. Stukenberg, P. T., Lustig, K. D., McGarry, T. J., King, R. W., Kuang, J., and Kirschner, M. W. (1997) *Curr. Biol.* **7,** 338–348.
28. Ferro, E. S., Tambourgi, D. V., Gobersztejn, F., Gomes, M. D., Sucupira, M., Armelin, M. C. S., Kipnis, T. L, and Camargo, A. C. M. (1993) *Biochem. Biophys. Res. Commun.* **191,** 275–281.
29. Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D., and Hill, C. P. (1997) *EMBO J.* **16,** 3787–3796.
30. Bode, W., Gomis-Rüth, F. X., and Stöckler, W. (1993) *FEBS Lett.* **331,** 134–140.
31. Lew, R. A., Hey, N. J., Tetaz, T. J., Glucksman, M. J., Roberts, J. L., and Smith, A. I. (1995) *Biochem. Biophys. Res. Commun.* **209,** 788–795.
32. Tisjlar, U., and Barrett, A. J. (1990) *FEBS Lett.* **264,** 84–86.
33. Tsukuba, T., and Bond, J. S. (1997) *J. Biol. Chem.* **273,** 35260–35267.
34. Murphy, G., and Knäuper, V. (1997) *Matrix Biol.* **15,** 511–518.
35. Valle, B. L., and Auld, D. S. (1990) *Biochemistry* **29,** 5647–5659.
36. Andersson, B., Wentland, M. A., Ricafrente, J. Y., Liu, W., and Gibbs, R. A. (1996) *Anal. Biochem.* **236,** 107–113.
37. Yu, W., Andersson, B., Worley, K. C., Muzny, D. M., Ding, Y., Liu, W., Ricafrente, J. Y., Wentland, M. A., Lennon, G., and Gibbs, R. A. (1997) *Genome Res.* **7,** 353–358.
38. Portaro, F. C. V., Gomes, M. D., Cabrera, A., Fernandes, B. L., Silva, C. L., Ferro, E. S., Juliano, L., and Camargo, A. C. M. (1999) *Biochem. Biophys. Res. Commun.* **255,** 596–601.
39. Silva, C. L., Portaro, F. C. V., Bonato, V. L. D., Camargo, A. C. M., and Ferro, E. S. (1999) *Biochem. Biophys. Res. Commun.* **255,** 590–595.

50 SIGMA®

NEW PRODUCTS
ALPHABETICAL LIST
BIOACTIVE PEPTIDES
IMMUNO-CHEMICALS
MOLECULAR BIOLOGY
TISSUE CULTURE
OTHER PRODUCT GROUPS
EQUIPMENT BOOKS AND SUPPLIES
DIAGNOSTIC KITS AND REAGENTS
PRODUCT INDEX



BIOCHEMICALS AND REAGENTS FOR LIFE SCIENCE RESEARCH



TO PLACE AN ORDER 800-325-3010
Fax: 800-325-5052
24 hours a day, seven days a week
Holidays: 6:00 a.m. to 5:00 p.m. (CST)
For large scale inquiries, see page 16.

CUSTOMER SERVICE 800-325-8070
To discuss the status of an existing order.

TECHNICAL SERVICE 800-325-5832
To discuss applications or product-specific questions.

GENERAL INFORMATION 800-521-8956
Should you have any problems getting excellent service at the above numbers, please call COLLECT 314-531-9500 anytime day or night, including weekends and holidays.

TIDES

US $

OBLAST GROWTH FACTOR, BASIC

agment 1-24 100 µg 28.35
rvine 250 µg 51.45
o-Ala-Leu-Pro-Glu-Asp-Gly-Gly-
r-Gly-Ala-Phe-Pro-Pro-Gly-His-Phe-Lys-Asp-Pro-Lys-
g-Leu-Tyr)
LC shows two peaks which are believed to be dif-
ent conformations of the same peptide.
f.: Esch, F., et al., Proc. Nat. Acad. Sci. USA, **82**,
07 (1985).
2031-54-3] FW 2553.9

agment 106-120 100 µg 18.40
rr-Arg-Ser-Arg-Lys-Tyr-Ser-Ser- 250 µg 40.45
o-Tyr-Val-Ala-Leu-Lys-Arg) 1 mg 131.25
etate Salt
inimum 95% (HPLC)
f.: Baird, A., et al., Proc. Nat. Acad. Sci. USA, **85**,
24 (1988).
14752-18-0] FW for free base 1963.3

ONECTIN FRAGMENTS and RELATED
EPTIDES
e:
g-Gly-Asp *Page 1116*
g-Gly-Asp-Ser *Page 1116*
g-Gly-Asp-Ser-Pro-Ala-Ser-Ser-Lys-Pro *Page 1117*
g-Gly-Glu-Ser *Page 1117*
g-Phe-Asp-Ser *Page 1117*
s-Gln-Asp-Ser-Glu-Thr-Arg-Thr-Phe-Tyr *Page 1117*
y-Arg-Ala-Asp-Ser-Pro-Lys *Page 1117*
y-Arg-Gly-Asp *Page 1118*
y-Arg-Gly-Asp-Ser *Page 1118*
y-Arg-Gly-Asp-Ser-Pro-Lys *Page 1118*
y-Arg-Gly-Asp-Thr-Pro *Page 1118*
r-Asp-Gly-Arg-Gly *Page 1118*

JCULAR GONADOTROPIN- 100 µg 43.55
ELEASING PEPTIDE 500 µg 138.95
man
r-Asp-Thr-Ser-His-His-Asp-Gln-Asp-His-Pro-Thr-Phe-
n)
inimum 95% (HPLC)
f.: Li, C.H., et al., Proc. Nat. Acad. Sci. USA, **84**,
9 (1987).
07873-08-5] FW 1651.6

ANIN (1-13)- 100 µg 33.10
BRADYKININ (2-9) AMIDE 500 µg 110.25
y-Trp-Thr-Leu-Asn-Ser-Ala-Gly-Tyr-
u-Leu-Gly-Pro-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-NH_2)
inimum 97% (HPLC)
ndem-linked peptide fragments produce high-
finity competitive galanin antagonist.
f.: 1. Oegren, S.O., et al., Neuroscience, **51**, 1
992).
Bartfai, T., et al., TIPS **13**, 312 (1992).
42846-71-7] FW 2233.6

ANIN (1-13)-SPANTIDE I 100 µg 51.00
y-Trp-Thr-Leu-Asn-Ser-Ala-Gly-Tyr-
u-Leu-Gly-Pro-o-Arg-Pro-Lys-Pro-Gln-Gln-o-Trp-Phe-
Trp-Leu-Leu-NH_2)
inimum 97% (HPLC)
ndem-linked peptide which acts as a potent galanin
ceptor antagonist.
f.: 1. Crawley, J.N., et al., Brain Res., **600**, 268
993).
43868-20-6] FW 2828.3

o use price list - page 2.

MISCELLANEOUS PEPTIDES

PRODUCT NUMBER			US $
G 9898 -0°C	**GLN-ALA-THR-VAL-GLY-ASP-ILE-ASN-THR-GLU-ARG-PRO-GLY-MET-LEU-ASP-PHE-THR-GLY-LYS** **Human** (Diazepam Binding Inhibitor [DBI] Fragment 51-70) **Minimum 95% (HPLC)** **Ref.:** Gray, P.W., et al., Proc. Nat. Acad. Sci. USA, **83**, 7547 (1986). *[104360-70-5]* FW 2150.4	500 µg 1 mg 5 mg	33.10 55.15 248.10
G 3642 -0°C	**GLN-ALA-THR-VAL-GLY-ASP-VAL-ASN-THR-ASP-ARG-PRO-GLY-LEU-LEU-ASP-LEU-LYS** (Octadecaneuropeptide; ODN) **Minimum 90% (HPLC)** A tryptic peptide of the diazepam binding inhibitor. Binds to the benzodiazepine receptor and appears to increase anxiety. **Ref.:** 1. Marx, J.L., Science, **227**, 934 (1985). 2. Ferrero, P., et al., Neuropharmacology, **23**, 1359 (1984). *[95237-86-8]* FW 1912.1	100 µg 250 µg 1 mg	19.65 28.35 78.75
G 9016 -0°C	**GLN-ARG-ARG-GLN-ARG-LYS-SER-ARG-ARG-THR-ILE** **Minimum 95% (HPLC)** **Peptide content: Approx. 65%** The C-terminal sequence of human interleukin-2 receptor; may be useful in the assay of protein kinase C. **Ref.:** Gallis, B., et al., J. Biol. Chem., **261**, 5075 (1986). *[102579-44-2]* FW 1484.7	100 µg 250 µg 500 µg	15.00 25.00 40.00
G 2637 -0°C	**GLU-ALA-GLU** (Thymosin α_1 Fragment 25-27) **Minimum 97% (HPLC)** **Ref.:** 1. Birr, C. and Stollenwerk, U., Angew. Chem. Int. Ed. Engl., **18**, 394 (1979). 2. Birr, C., et al., Peptides, Synthesis–Structure–Function, Proc. 7th Amer. Pept. Symp., Rich, D. and Gross, E., eds., 541 (1981). *[6234-27-1]* $C_{14}H_{21}N_3O_8$ FW 347.3	1 mg 5 mg 10 mg	16.20 51.80 85.70
G 5775 -0°C	**GLU-ALA-GLU-ASN** (Thymosin α_1 Fragment 25-28) **Ammonium Salt** **Minimum 95% (HPLC)** **Ref.:** 1. Birr, C. and Stollenwerk, U., Angew. Chem. Int. Ed. Engl., **18**, 394 (1979). 2. Birr, C., et al., Peptides, Synthesis–Structure–Function, Proc. 7th Amer. Pept. Symp., Rich, D. and Gross, E. eds., 541 (1981). *[78603-79-9]* $C_{17}H_{27}N_5O_{10}$ FW 461.4 (for free acid)	1 mg 5 mg 10 mg	16.15 52.30 85.95
G 3765 -0°C	**γ-GLU-GLY** **Minimum 97% (HPLC)** Excitatory amino acid antagonist. **Ref.:** Collingridge, G.L., et al., J. Physiol., **334**, 19 (1983). *[6729-55-1]* $C_7H_{12}N_2O_5$ FW 204.2	25 mg 100 mg 250 mg	15.75 36.75 68.25
G 1153 -0°C	**GLU-SER-PRO-LEU-ILE-ALA-LYS-VAL-LEU-THR-THR-GLU-PRO-PRO-ILE-ILE-THR-PRO-VAL-ARG-ARG** (Phospholipase A_2 activating peptide) **Minimum 95% (HPLC)** **Peptide content: Approx. 70%** **Ref.:** Clark, M.A., et al., Proc. Nat. Acad. Sci. USA, **88**, 5418 (1991). *[137314-60-4]* FW 2330.8	100 µg 500 µg	24.85 82.70
G 5149 -0°C	**GLY-GLN** **Hydrate** **Minimum 97% (TLC)** An inhibitory neuropeptide.[1-3] A heat-stable substitute for glutamine in mammalian cell culture media.[4] **Ref.:** 1. Parish, D.C. and Smyth, D.G., Biochem. Soc. Trans, **10**, 221 (1982). 2. Parish, D.C., et al., Nature, **306**, 267 (1983). 3. Koelle, G.B., et al., Proc. Nat. Acad. Sci. USA, **82**, 5213 (1985). 4. Roth, E., et al., In Vitro Cell. Dev. Biol., **24**, 696 (1988). *[13115-71-4]* $C_7H_{13}N_3O_4$ FW 203.2	1 g 5 g	18.90 63.00
G 3774 -0°C	**GLY-GLU-GLN-ARG-LYS-ASP-VAL-TYR-VAL-GLN-LEU-TYR-LEU** (Thymopoietin II Fragment 29-41) **Minimum 97% (HPLC)** Domain containing the active site of thymopoietin II (fragment 32-36). **Ref.:** 1. Schlesinger, D.H., et al., Cell, **5**, 367 (1975). 2. Goldstein, G., et al., Science, **204**, 1309 (1979). *See also: Arg-Lys-Asp-Val-Tyr Page 1148* *[56795-64-3]* FW 1610.8	1 mg 5 mg	41.20 154.35
G 6887 -0°C	**GLY-GLY-ARG** **Acetate Salt** **Minimum 97% (HPLC)** **Peptide content: Approx. 70%** *[55033-48-2]* $C_{10}H_{20}N_6O_4$ FW 288.3 (for free base)	1 mg 5 mg 10 mg 50 mg	12.05 33.10 50.75 144.25
G 5772 -0°C	**GLY-GLY-HIS** (Copper binding peptide) **Acetate Salt** **Minimum 97% (TLC)** **Ref.:** 1. Lau, S.J., et al., J. Biol. Chem., **249**, 5878 (1974). 2. Iyer, S., et al., Biochem. J., **169**, 61 (1978). *[93404-95-6]* $C_{10}H_{15}N_5O_4$ FW 269.3 (for free base)	100 mg 250 mg 500 mg	29.65 59.30 106.80
G 5386 -0°C	**GLY-GLY-TYR-ARG** **Minimum 97% (HPLC)** Inhibitor of papain Useful for affinity chromatography **Ref.:** Funk, M.O., et al., Int. J. Pept. Prot. Res., **13**, 296 (1979). *[70195-20-9]* $C_{19}H_{29}N_7O_6$ FW 451.5	5 mg 25 mg 100 mg	5.80 16.30 60.90
G 8636 -0°C	**GLY-HIS-ARG-PRO** (Human fibrin β-chain fragment 1-4) **Acetate Salt** **Minimum 97% (HPLC)** **Ref.:** 1. Laudano, A.P. and Doolittle, R.F., Proc. Nat. Acad. Sci. USA, **75**, 3085 (1978). 2. Laudano, A.P. and Doolittle, R.F., Biochemistry, **19**, 1013 (1980). *[103213-28-1]* $C_{19}H_{31}N_9O_5$ FW 465.5 (for free base)	1 mg 5 mg 10 mg 25 mg	12.15 31.50 64.50 102.90
G 1887 -0°C	**GLY-HIS-LYS** (Liver cell growth factor) **Acetate Salt** **Minimum 97% (TLC)** **Peptide content: Approx. 70%** **Ref.:** 1. Pickart, L., et al., Biochem. Biophys. Res. Commun., **54**, 562 (1973). 2. Pickart, L., et al., J. Chromat., **175**, 65 (1979). *[72957-37-0]* $C_{14}H_{24}N_6O_4$ FW 340.4 (for free base)	5 mg 10 mg 50 mg	14.25 23.70 94.75
G 5275 -0°C	**GLY-PEN-GLY-ARG-GLY-ASP-SER-PRO-CYS-ALA** [Disulfide Bridge: 2—9] **Minimum 97% (HPLC)** *[111844-17-8]* $C_{35}H_{57}N_{13}O_{14}S_2$ FW 948.0	100 µg 500 µg 1 mg	14.20 47.25 78.75